TOWARDS STRUCTURAL STUDIES OF TRANSCRIPTIONALLY-COUPLED CRISPR SYSTEMS

by
Bobby Kwan

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland
April 2017

ABSTRACT

Bacteria and archaea employ an adaptive immune system in the form of clustered regularly interspaced short palindromic repeats (CRISPR) to fend off mobile genetic elements. CRISPR arrays store immunological memory in the form of spacer sequences flanked by repeat sequences, where the array is preceded by a leader sequence to regulate the placement of acquired protospacers. CRISPR-associated (Cas) proteins generate and carry out the immune response against invading genetic elements.

Categorized as a Class I CRISPR system, Type III systems are further divided into Type III-A (Csm) and III-B (Cmr) which utilize multisubunit complexes to target and cleave DNA. The Cmr complex binds and cleaves RNA transcripts to activate its DNA cleavage activity. The goal of obtaining a cryo-EM structure of the Cmr complex is to determine how target RNA binding activates DNase activity, as the target binding site and the DNase active site are distant from each other.

The Thermotoga maritima Cmr2 – 6 complex was purified by pooling together cells expressing each subunit and lysing by microfluidizer, followed by immobilized metal ion affinity (IMAC) and size exclusion chromatography. The Cmr2:3 heterodimer was purified in a similar manner, with additional desalting and polyhistidine (His) tag cleavage steps after IMAC elution. Purification of the Cmr complex produced 180µg of protein, and purification of Cmr2:3 yielded 7mg of protein. Crystallization trays were set up with the purified Cmr2:3 sample, and the Cmr2 – 6 complex was incubated with Cmr1 and CRISPR RNA to form the full interference complex. 500µm crystal florets formed in the Cmr2:3 crystallization trays, but x-ray diffraction determined that Cmr3 alone formed the crystals. The Cmr complex incubation mixture was separated by analytical size

exclusion, and the monodisperse peak corresponding to the correct number of subunits was recovered for cryo-EM analysis.

The disruption in Cmr2:3 crystallization was thought to be a result of the His tag interfering with the crystal packing of the heterodimer, and new crystallization trays were set up with tag-cleaved Cmr2:3. The full Cmr complex was also successfully formed, but the bound RNA needs to be analyzed to determine whether it is intact after full complex formation.

Advisor: Scott Bailey

Secondary Reader: Jennifer Kavran

ACKNOWLEDGEMENTS

I would like to first and foremost thank my parents for leaving everything behind in China to find new opportunities in America. They have provided opportunities for me that they did not have themselves, and I owe most of my achievements thus far to them for their support. I also want to thank my brother Ryan for defending and supporting me in times of doubt and conflict, and for listening to me when I had no others to confide in.

I want to also thank Janice Evans, Shannon Gaston, and Sean Prigge for being the first people in the Johns Hopkins community to let me pitch my story in the hopes of getting admitted into the MHS program. They were the ones who made the difficult decision to give me a chance and admit me, and I am forever indebted to them for this amazing opportunity to pursue my graduate education at Johns Hopkins University.

I want to thank my mentor and principal investigator Scott Bailey for always looking out for me throughout my entire masters career, for always pushing me when I needed it, for making sure that I meet the expectations of a rigorous PhD student, and for supporting me in my quest for admission into a PhD program no matter how improbable my chances were. I also want to thank John Mallon, Kaitlin Johnson, Anita Ramachandran, and Katelyn Jackson in the Bailey lab for being a major source of support and insight while I conducted research for my thesis, and for supporting me in all my academic endeavors.

Lastly, I would like to thank my friends Anna, Jasmine, Brandi, Garrett, Susan, and Molly for picking me up when I was down, for supporting me when I was least confident in myself, and for cheering me up when I needed it the most.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Humans utilize an immune system to fend off foreign infections by viruses and bacteria using an innate immune response upon initial infection, and a memory-based adaptive immune response upon subsequent infections by the same microbe. 90% of sequenced archaea and 40% of sequenced bacteria also employ a similar type of immune system against mobile genetic elements using clustered regularly interspaced short palindromic repeats (CRISPR). CRISPR arrays store immunological memory in the form of spacer sequences flanked by repeat sequences, where the array is preceded by a leader sequence to regulate the placement of acquired protospacers from mobile genetic elements. CRISPR-associated (Cas) proteins generate and carry out the adaptive immune response against invading genetic elements.

CRISPR Components

CRISPR immunity is subdivided into two classes, composed of numerous subtypes each with its own distinct interference mechanism and molecular composition. Interference is the mechanism of CRISPR response where ribonucleoprotein single/multi-subunit complexes composed of Cas proteins surveil the intracellular environment guided by bound CRISPR RNA (crRNA) and cleaves foreign DNA that anneal to the crRNA. Class I CRISPR systems utilize a multi-subunit interference complex composed of numerous Cas proteins, crRNA, and includes Type I, III, and IV systems (Figure 1). Class II CRISPR systems utilize a single subunit interference complex composed of one

Cas protein with crRNA, held together by a structural transactivating CRISPR RNA (tracrRNA) in the Type II subtype. Protospacers are the sequences of foreign DNA that are excised by Cas proteins and integrated into the host CRISPR array, and therefore are the target sequences for CRISPR interference. CRISPR RNA is composed of a single spacer sequence flanked by truncated repeat sequences, which is used as the guide to selectively target mobile genetic elements that forms base complementarity to the protospacer sequence. Type I, II, and V systems utilize short 2-5 nucleotide sequences that flank the protospacer in foreign DNA called protospacer adjacent motifs (PAMs) as markers for distinguishing between self and nonself DNA (Wang et al., 2015). During acquisition of spacer sequences, spacers are preferentially selected from protospacers that contain a cognate PAM to that CRISPR system(Mojica et al., 2009). Other CRISPR types use other mechanisms to avoid autoimmunity, such as the Type III systems utilizing the flanking sequences of the RNA transcript to determine whether the DNase activity of Cas10-like Cmr2 will be turned on or off (Marraffini and Sontheimer, 2010). Acquisition is a process that involves the excision of protospacers from foreign DNA and the integration of the protospacers into the host CRISPR array. Acquisition components are conserved across all classes and subtypes, and features Cas1, Cas2, the CRISPR array leader sequence, and the first repeat sequence after the leader sequence. Cas1 exhibits nuclease activity and inserts new spacers in the CRISPR array between the leader sequence and the first repeat, with the leader sequence being a long A:T-rich sequence oriented immediately upstream of the CRISPR array, containing both the crRNA promoter and the recognition sequence for spacer insertion (Diez-Villasenor et al., 2013).

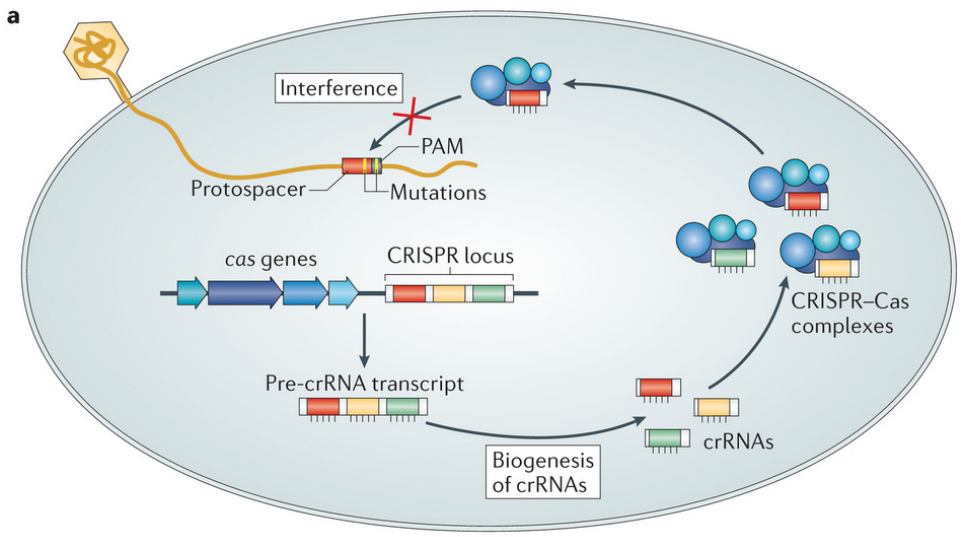


Figure 1: Pathway of Class I CRISPR system response (Samson et al., 2013)

Mechanism of Acquisition

Upon invasion by foreign DNA, the host CRISPR system needs to distinguish the incoming DNA as self vs nonself before acquisition is initiated. A sequence of 30-40 base pairs needs to be obtained from the foreign DNA, which will serve as a protospacer for insertion into the host CRISPR array. Two types of adaptation have been determined in Type I systems, the first of which is naïve acquisition which takes place upon encounter with a new invading genetic element, whereas primed acquisition occurs if the invading phage mutates their PAM or protospacer sequence in attempts to evade interference (Datsenko et al., 2012). Naïve adaptation utilizes only Cas1 and Cas2, whereas primed adaptation also requires the Cascade interference complex as well as the Cas3 endonuclease. Cas3 is composed of a histidine-aspartate (HD) domain that exhibits DNase activity, which is fused to a helicase domain that unwinds dsDNA in an ATP-dependent manner. The necessity of Cas1 in general spacer acquisition was confirmed by Babu et al. by generating the nuclease-deficient Cas1 D221A mutant and demonstrating that the construct did not support spacer acquisition *in vivo* (Babu et al., 2011). Naïve adaptation in Type II systems however require Cas9 to recognize PAMs in comparison to Type I-E systems where only Cas1 and Cas2 are needed for detection of PAMs. Primed adaptation on the other hand does not occur unless Cas1, Cas2, Cas3, Cascade, and the priming spacer are present (Datsenko et al., 2012). Type I-E systems utilize priming to amplify acquisition ten- to twenty-fold over naïve adaptation, and is biased towards the strand orientation that matches the protospacer that is targeted by the priming spacer (Savitskaya et al., 2013). Datsenko et al. postulated that priming likely evolved as a

mechanism of avoiding infection by phage escape mutants that would otherwise evade interference by CRISPR, as priming efficiency is markedly increased when the priming spacer was mutated in the seed sequence or if the protospacer contained a non-cognate PAM (Datsenko et al., 2012). Single-molecule experiments have also shown that non-canonical Cascade binding modes exist at mutated protospacers and could be crucial to the primed adaptation mechanism as protospacers with mutated PAMs can still elicit stable Cascade binding. These results coupled with the identification of specific intermolecular Cascade:Cas3 interactions shows that recognition of a consensus PAM by Cascade is necessary for functional recruitment of Cas3 to promote interference (Blosser et al., 2015).

RecBCD: Protospacer Source

RecBCD is an exonuclease complex that recognizes exposed dsDNA from double-stranded breaks in *E. coli*, which it then unwinds and degrades until it reaches an 8bp crossover hotspot instigator (Chi) site. Hotspots for protospacers are situated between the stall sites of replication forks, a major source of double-stranded breaks, and the closest Chi site (Ivancic-Bace et al., 2015). Chi sites appear frequently in the *E. coli* genome, typically occurring every 4.6 kilobases (kb) compared to 65kb if the sites were to appear by chance in a genome. The frequency of Chi sites allows for only short lengths of host DNA to be degraded by RecBCD before it reaches the next Chi site, which yields very few potential spacers that arise from degradation products of host DNA (Levy et al., 2015). This observation when applied to foreign DNA that are not enriched in Chi sites

will result in longer stretches of DNA degradation by RecBCD, yielding a greater number of viable protospacers for integration. Since phage/viral DNA enter hosts as linear dsDNA, the linear ends are recognized by RecBCD as double-stranded breaks which promotes the degradation of the foreign DNA until RecBCD reaches the next Chi site in the sequence. The Cas1:2 heterohexamer along with RecBCD obtain the protospacer from the foreign DNA and integrates it as a spacer in the CRISPR array after duplication of the flanking repeat sequence (Nunez et al., 2014). Testing by Yosef et al. of repeats by sequencing of the duplicated repeat showed that the newly inserted repeat is identical to the repeat that is proximal to the leader sequence. Spacers are added to the array in a polar manner, orienting towards the leader sequence side of the array where the most recently-incorporated spacer is proximal to the leader sequence (Yosef et al., 2012). *E. coli* protospacers containing a flanking 5'-AAG-3' PAM are selected as donors with a 35% frequency, which is overrepresented compared to the frequency of PAMs in general, which is roughly 1.6%. Protospacers of varying GC content did not appreciably affect the efficiency of acquisition, but an AA motif at the end of the protospacer did increase the efficiency of adaptation (Fineran et al., 2014).

Acquisition Mechanism: Adaptation

Mechanistically, naïve adaptation requires DNA substrates that contain the 3'-OH that is needed for nucleophilic attack on both strands of the repeat. The Cas1:2 heterohexamer integrates new spacers with the correct orientation of the PAM by

utilizing the 3'-OH cytidine that is complimentary to the third residue of the AAG PAM (Pougach et al., 2010). Regulation of adaptation is crucial to prevent the acquisition of self spacers that would lead to cell death, and thus transcription-dependent regulators of CRISPR interference exists in several systems such as the *E. coli* Type I-E cyclic AMP receptor protein regulator and the histone-like nucleoid-structuring protein (H-NS) (Pougach et al., 2010). H-NS suppresses the operon promoter that is composed of the Cascade, Cas1, Cas2, and transcription regulation genes of the CRISPR array. Because of this, H-NS reduces adaptation by repressing the promoter of the cas gene operon. The redundancy in PAM recognition serves to enhance the specificity of spacer acquisition using consensus PAMs during primed adaptation over naïve adaptation (Savitskaya et al., 2013). Experiments on naïve adaptation have shown that Cas1 and Cas2 of Type I-E systems prefer spacers from plasmids as opposed to chromosomal DNA, despite the excess of chromosomal DNA available to Cas1 and Cas2, which is due to the inherent selectivity for foreign DNA during the adaptation process as opposed to selective pressures against acquisition of chromosomal spacers (Yosef et al., 2012). A major mechanism of preference for foreign DNA reveals spacer acquisition hotspots upstream of replication stalling (Ter) sites, which stall the faster-moving replication forks during bidirectional DNA replication until the lagging replication fork finishes replication to allow for chromosome decatenation (Levy et al., 2015). Additional hotspots for spacer acquisition include DNA nicks and double-stranded breaks that also stall replication forks, which are targeted by RecBCD for degradation until it reaches the next Chi site. This suggests that spacer acquisition occurs most frequently from regions of DNA that

undergo frequent replication stalls, and the high copy number of plasmids results in a greater occurrence of termination events in comparison to chromosomal replication.

Structural Mechanism of Cas1-Cas2 Function

Structurally, interactions between N-terminal domains of adjacent Cas1 protomers promotes stable homodimerization of Cas1 (Nunez et al., 2015). Three highly-conserved C-terminal residues (E141, H208, and D221 in *E. coli*) coordinate divalent metal ions to form the putative active site of Cas1, and mutation of any of these residues results in complete elimination of observable nuclease activity. Cas2 possesses a highly-conserved architecture composed of a single domain and a ferredoxin-like fold that is also found in Cas5 and Cas6, and readily forms a stable homodimer like in Cas1 (Nunez et al., 2015). *E. coli* Cas1:2 forms a heterohexamer architecture with pseudo-two-fold symmetry containing a central Cas2 dimer that interacts on two opposing faces with separate Cas1 dimers. The Cas1:2 interface is stabilized by electrostatic and hydrophobic interactions, and interference of this interface results in perturbation of complex formation *in vitro* and spacer acquisition *in vivo*. Binding between CRISPR DNA and Cas1 is dependent upon the presence of Cas2, but acquisition was completely inhibited by Cas1 active site mutations (Nunez et al., 2014). This suggests that the primary role of Cas2 involves the repositioning of Cas1 in the heterohexamer complex as opposed to exhibiting specific nuclease activity. The Cas1:2 heterohexamer binds splayed dual-fork DNA substrates where 23 bases of dsDNA are flanked by 3' single-stranded overhangs that thread into two Cas1 active sites. Tyrosine-22 of the Cas1 monomers brace the double-stranded

region via stacking interactions, which defines the "caliper" role of Cas1:2 to accurately measure the length of new spacers, as well as functioning as wedges to generate junctions between single-stranded and double-stranded regions of DNA (Wang et al., 2015). Cas1 specifically targets the 5'-CTT-3' PAM complementarity sequence within the 3' single-stranded overhang region of a donor spacer, which helps position the phosphodiester bond after the C1 base to the active site of Cas1 to allow for trimming of spacer precursors to the correct length.

Mechanism of Guide RNA Biogenesis

To determine interference targets, CRISPR systems require guides which are transcribed from CRISPR arrays. Upon addition of protospacers into the CRISPR array, the array is then transcribed as a precursor crRNA (pre-crRNA) containing all the spacers in the array. Depending on the system subtype, the pre-crRNA is matured in one or two stages; in Type I and III systems, a Cas6-family endoribonuclease cleaves the pre-crRNA within the repeat regions to yield single spacers flanked by incomplete repeat sequences (Carte et al., 2010). Cas6 nucleases cleave the pre-crRNA by hydrolyzing an individual phosphodiester bond in the repeats of transcripts, resulting in crRNAs that contain a complete spacer flanked by an 8 nucleotide 5' handle derived from repeats, and a 3' handle of variable length derived from repeats as well (Carte et al., 2008).

Cas6-mediated RNA Processing

Of the three known processing endonucleases Cas6, Cas6e, and Cas6f, each share very little sequence homology but instead share similar three-dimensional structures that exhibit ferredoxin folds as RNA recognition motifs (RRMs), where RNA-bound Cas6f contains a single ferredoxin fold while Cas6e and Cas6 both contain tandem ferredoxin folds. Biochemically, Cas6 homologs are metal-independent endoribonucleases that cleave pre-crRNA to generate crRNAs that have a 5' hydroxyl and either a cyclic 2'-3' phosphate or a 3' phosphate (Wang et al., 2011). Of the identified CRISPR endoribonucleases, they all share the same nucleolytic activity of cleaving 8 nucleotides upstream of the repeat sequence junction. CRISPR repeat RNA that are cleaved by *Pseudomonas aeruginosa* Cas6f contain palindromic sequences that form stable hairpin loops that are directly upstream of the cleavage site. Cas6f then interacts specifically with the RNA hairpin to position the cleavage site at the base of the hairpin, which is also demonstrated in *E. coli* Cas6e (Brouns et al., 2008) but not in *Pyrococcus furiosus* Cas6 as it clamps onto nucleotides 2-9 of the repeat RNA through its two ferredoxin-like folds (Kunin et al., 2007). The Cas6-anchored 5' end of the repeat RNA tethers the cleavage site between nucleotides 22 and 23 to the Cas6 active site (Wang et al., 2011). While sequences proximal to the cleavage site are important for Cas6 cleavage activity, the distal sequences near the 5' terminus of the repeat RNA are also necessary for Cas6 binding (Figure 2). A notable finding is that recognition of CRISPR RNAs by Cas6 occurs at a site that is distal to the crRNA cleavage site, and superimposition of the RNA-bound and RNA-free Cas6 structures shows that helix α2 containing the catalytic His46

residue is rotated 6° toward helix α1 in the RNA-bound structure, and this shift demonstrates that RNA binding induces changes in the Cas6 catalytic region positioned at the opposite side of the protein. Tyrosine-31, histidine-46, and lysine-52 compose the catalytic triad of *P. furiosus* Cas6, as substitution mutations results in inhibition of detectable RNAse activity without hindering RNA binding (Carte et al., 2010). Substitution of CRISPR repeat RNA nucleotides 25-30 downstream of the cleavage site does not interrupt binding by Cas6 but will prevent cleavage as nucleotides U25 and G26 are crucial for cleavage. The unobserved region of repeat RNA between nucleotide 10-22 that binds to the crRNA-recognition interface of Cas6 is lined with positive electrostatic potential and passes over G-rich loop that is specific to Cas6. Substitution mutations of glycine 223, 225, 231, and 233 to alanine resulted in complete inhibition of RNAse activity while preserving weak binding activity to RNA in Cas6 (Wang et al., 2011).

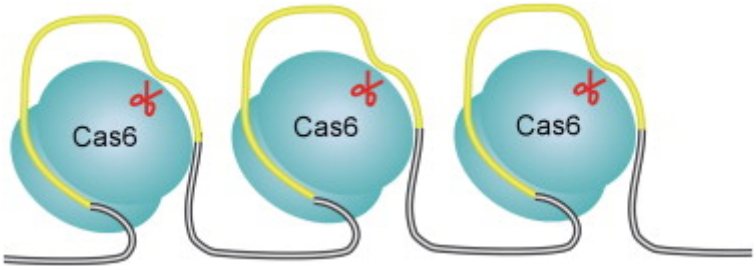


Figure 2: beads-on-a-string model of crRNA binding to Cas6, with 5' terminus of crRNA needed for wrap-around binding (Wang et al., 2011)

CRISPR RNA Processing

In Type II systems, the pre-crRNA base pairs with tracrRNA to form an RNA heteroduplex which is then processed by RNase III in the presence of Cas9 (Deltcheva et al., 2011). Upon transcription and maturation of the guide RNAs, they are then associated with their single- or multi-subunit CRISPR interference complex depending on the system subtype. After the crRNA are transferred to the interference complexes, the 3' handles of the crRNA become accessible for trimming by nucleases. In Type I-A, I-B, I-C, and I-D systems, the product of nucleolytic trimming contains shorter 3' handles, whereas differential trimming in Type III systems yield two mature crRNAs that differ by 6 bases (Staals et al., 2014).

Mechanism of Interference

In Type I systems, interference requires recognition of a PAM sequence in the invading DNA as well as annealing to the crRNA of the multisubunit Cascade interference complex to form an R-loop composed of the double-stranded DNA target hybridizing with the single-stranded crRNA, forming a loop from the displaced non-complementary DNA strand (Westra et al., 2012). Upon satisfying these criteria, Cascade then undergoes a conformational change to recruit the Cas3 endonuclease for complete cleavage of the invading DNA in the 3' to 5' direction (Sinkunas et al., 2011). For Type I systems to successfully interfere foreign DNA, base pairing between the seed sequence of crRNA and the complementary protospacer is crucial at positions most proximal to the PAM, which constitutes nucleotides 1-5 and 7-8 at the 5' end of the crRNA in Type I-E

systems (Semenova et al., 2011). Incomplete base pairing with the seed sequence, as demonstrated by mutant targets, results in evasion of Cascade interference. Upon PAM recognition and complete annealing of the crRNA seed sequence to the complementary DNA target, dsDNA cleavage occurs in an ATP-dependent manner.

Type II systems utilize Cas9 in complex with tracrRNA and crRNA to interfere foreign DNA in a PAM-dependent manner through its RuvC-like and HNH endonuclease domains (Gasiunas et al., 2012). The recognition of PAMs in Cas9 is governed by two C-terminal tryptophan-containing loops of the nuclease domain, and substitution of these residues negatively affects binding and cleavage of target DNA. Finding protospacer sequences that are complementary to the crRNA involves surveillance of the foreign DNA and annealing at a 7-8 nucleotide seed sequence of the crRNA and the complementary protospacer, which promotes extended base pairing along the rest of the crRNA sequence and causes an R-loop structure to form which results in a conformational change to the interference complex to activate intrinsic nuclease activity in Type II and III-B systems (Reeks et al., 2013).

In Type III systems, the interference complex is similar to Cascade in Type I but it does not exclusively target foreign DNA as it also targets phage mRNA as well, where processive cleavage of the RNA target occurs in 6 nucleotide intervals in the 3' to 5' direction (Deng et al., 2013). The RNase activity of the Cmr4 subunits is implicated in the regularity of cleavage patterns of Cmr interference complexes. *Pyrococcus furiosus* type III-B Cmr complexes re-formed with single crRNAs will cleave target RNAs in specific six nucleotide intervals within the region of crRNA complementarity; with four

cleavage events occurring using 45-nucleotide crRNAs and three cleavage events occurring using 39-nucleotide crRNAs (Hale et al., 2014). Efficiency of target RNA cleavage decreases when the crRNA length is reduced to 20 nucleotides, because of reduced RNA hybridization between the crRNA and the target RNA or from less efficient Cmr complex formation. As crRNA length is shortened from the 3' terminus, the number of cleavage events occurring at the 5' terminus of the target RNA are gradually lost (Hale et al., 2014). The number of cleavage events performed by the Cmr complex can be increased by elongating the guide sequence; extending the 45-nucleotide crRNA guide region by 22 nucleotides results in four additional cleavage events occurring on the target RNA at regular 6 nucleotide intervals.

RNA cleavage activity of *Streptococcus thermophilus* Type III-A Csm functions similarly to Cmr in that cleavage occurs at 6 nucleotide intervals through the multimeric Csm3 subunit, as well as requiring divalent metal cations to exhibit RNase activity. *S. thermophilus* Csm3 contains a well-conserved RNA recognition motif at its core and the conserved aspartate-33 residue has been confirmed through mutagenesis to be a critical element in the RNA cleavage activity of Csm. Although *S. thermophilus* Csm requires target RNA sequence complementarity for cleavage to occur, there is no detriment to cleavage activity if unpaired flanking sequences or PAMs are included in the protospacer, as well as some tolerance for contiguous mismatches or end truncations (Tamulaitis et al., 2014).

Type III CRISPR Systems

Categorized as a Class I CRISPR system, Type III systems are further divided into Type III-A and III-B which utilize the multisubunit Csm (Type III-A) and the Cmr (Type III-B) complexes to bind and cleave DNA. The six types of subunits in Type III systems share structural homology (Figure 3) with Cmr1 having no homologs in Csm. There are no current complete structures of the Csm complex in any organism, but there are cryo-EM structures of the *Thermus thermophilus* Cmr complex by the Doudna group (Taylor et al., 2015) and crystal structures of hybrid *Pyrococcus furiosus/Archaeoglobus fulgidus* Cmr complex (Osawa et al., 2015). All published structures of the Cmr complex contain truncated Cmr2 subunits that are missing the HD domain that mediates DNase activity of the complex as it was found that the domain was not necessary for crystallization of the complex, so there are no current published structures of the intact Cmr complex in any studied organism (Figure 4). The HD domain exhibits the DNase activity of the complex but the DNA binding site is distal to the HD domain, and little is known how the binding of DNA activates the DNase activity of Cmr. The Doudna group have demonstrated that binding of the DNA target induces an overall conformational change in the *Thermus thermophilus* Cmr complex, and this finding contributes to the search for a structural mechanism for DNase activation upon binding of DNA (Taylor et al., 2015).

Subunit	Type III-A	MW (kD)	Gene Length (bp)	Type III-B	MW (kD)	Gene Length (bp)
5' Large Subunit	Csm1	83	2100	Cmr2	93.5	2400
5' crRNA Capping Subunit	Csm4	33	876	Cmr3	37.9	990
Major Filamenting Subunit	Csm3	27.5	741	Cmr4	32.8	864
Minor Filamenting Subunit	Csm2	16.7	420	Cmr5	14.6	360
3' Terminal Capping Subunit	Csm5	44	1100	Cmr1	50.2	1326
3'-crRNA-associated Subunit	Csm3	27.5	741	Cmr6	27.9	720

Figure 3: Structural homology of Cmr and Csm subunits

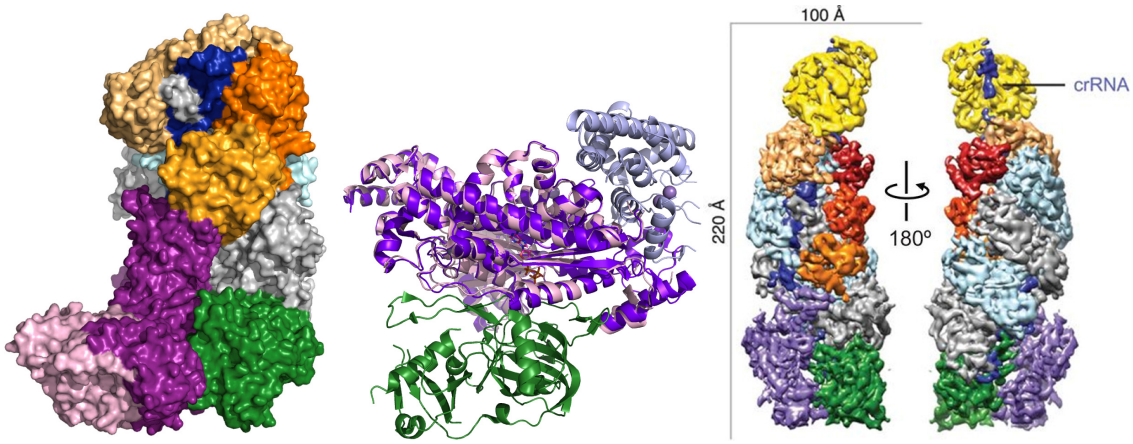


Figure 4: (From left to right) Left: model of Pfu Cmr complex (PDB ID: 3X1L) aligned with full-length Cmr2 (dark purple, PDB ID: 4W8Y) with HD domain (light pink). Center: Pfu Cmr2 full length (dark and light purple) superimposed with truncated Pfu Cmr2 (pink) in complex with Cmr3 (green) (PDB ID: 4H4K). Right: published cryo-EM reconstruction of Tth apo-Cmr complex with bound crRNA at 4.1Å (Taylor et al., 2015)

The Type III interference complexes bind and cleave RNA transcripts to activate their DNA cleavage activity (Zhang et al., 2012). The Csm and Cmr complexes binds and cleaves ssRNA targets in a divalent metal ion-dependent manner, and they both target DNA in a transcription-dependent manner *in vivo* (Samai et al., 2015). The architecture of Type III interference complexes are structurally similar in distribution, with the Type III-B Cmr2 and Cmr3 heterodimer (Csm1 and Csm4 in Type III-A systems) situated at the 5' handle of the crRNA. Four subunits of Cmr4 align along the sequence of the crRNA, whereas there are six subunits of Csm3 that are positioned along the crRNA, and three subunits of either Cmr5/Csm2 face opposite of the multimeric Cmr4/Csm3 (Taylor et al., 2015). Cmr1 and Cmr6 associate with the 3' terminus of the crRNA, whereas in Csm complexes there is only a single Csm5 associated with the 3' terminus. Cmr1 and Cmr6 form a cap on the 3' terminus of the complexed crRNA to mediate crucial interactions with the 5' terminus of the RNA substrate (Tamulaitis et al., 2014). The primary DNase activity is exhibited in the large subunits of the interference complexes, which are Csm1 and Cmr2, both featuring an HD domain.

Transcription is required to enter the lysogenic cycle of phage infection, and if transcription occurs through a protospacer region the nascent phage mRNA emerging from the transcription complex is targeted by Csm or Cmr (Goldberg et al., 2014). In order to avoid autoimmunity, measures must be taken to avoid targeting host DNA such as PAM targeting in Type I systems. Type III interference complexes are transcriptionally-coupled, and thus a mechanism of distinguishing self vs nonself targets

is required as the transcription of host anti-CRISPR sequences need to evade interference to avoid autoimmunity (Marraffini et al., 2010).

Formation of a Type III-B ribonucleoprotein complex with crRNA can occur in the absence of Cmr1 and Cmr6, but Cmr2-5 are all necessary for complete binding of the crRNA (Hale et al., 2014). UV light-induced RNA-protein crosslinking experiments have also demonstrated that the crRNA-binding interference complex formed with only Cmr2-5 possesses an organization that is identical to the full Cmr1-6 complex. The crosslinking experiments have also found that Cmr2, Cmr3, and Cmr4 were crosslinked to the fully-labeled crRNA, demonstrating that these subunits make direct contacts with the crRNA in the complex (Hale et al., 2014). Cmr3 crosslinks formed with the nucleotides in the 5' tag of the full-length crRNA as well as the eight nucleotide RNA tag itself without the guide sequence or 3' tail, whereas Cmr2 and Cmr4 only crosslinked to guide sequence nucleotides. Cmr2 crosslinking with the 5'-proximal region of the crRNA occurs alongside Cmr4 crosslinking along the span of the guide sequence that results in a reduction of crosslinking signal at the 3' terminus of the crRNA. While Cmr5 does not form any crosslink patterns with the crRNA, it is still necessary for the formation of the RNP complex with the other three necessary Cmr subunits (Spilman et al., 2013).

Native gel mobility shift assays have demonstrated that Cmr complex formation requires the 5'-OH and the 5' tag sequence of the crRNA, where substitution or deletion of the 5' tag sequence impairing binding of the crRNA by Cmr2-5. Substituting either all eight nucleotides or the first two nucleotides of the 5' tag sequence prevents assembly of the Cmr interference complex, and this is also seen in single nucleotide additions to the 5'

end of the tag as well as altering of the 5'-OH to a 5'-phosphate group (Hale et al., 2014). On the other hand, mutation of the 3' terminal sequence of the crRNA from the *in vivo* 3'-phosphate to a 3'-OH did not appreciably influence Cmr2-5 binding to the crRNA. Synthesized RNAs composed of only eight nucleotide repeats of the 5' tag sequence forms the Cmr complex with identical affinity compared to wild type full-length crRNA, with Cmr3 being the only subunit making direct contacts with the 5' tag sequence in absence and in the presence of the guide region of the crRNA. This indicates that Cmr1 and Cmr6 are crucial for target RNA binding, where elimination of any of the six subunits of a full RNA-binding Cmr interference complex results in significant inhibition of observed binding to the target RNA. Cmr3 identifies the crRNA sequence motif on its own, but requires protein-protein interactions within a complex that includes Cmr2, Cmr4, and Cmr5 for optimal binding activity to the crRNA (Taylor et al., 2015).

Cmr3 of the Cas5 superfamily makes direct contacts with the 5' terminal tag of crRNAs and plays a crucial role in crRNA recognition, as demonstrated from crystal structures of the Cas5 homolog in Type I-E interference complexes containing bound crRNA exhibiting base-specific contacts between the 5' tag and residues of Cas5 (Shao et al., 2013). This phenomenon play an important and well-conserved role in crRNA recognition and complexing in Type I and III systems. A large fraction of processed RNAs found in prokaryotes feature a 5'-phosphate termini, but Cas6 CRISPR transcript cleavage products include 5'-OH groups instead which are a distinguishing feature of crRNA from other processed prokaryotic RNA (Carte et al., 2008). The 5'-OH end of processed crRNAs are crucial for the successful discrimination between incorporation of

crRNA and the potentially lethal incorporation of host RNAs onto CRISPR interference complexes.

CRISPR in Public Health: Drug Discovery

CRISPR technology is lauded for its precision and applicability in excision/insertion of genetic elements across a vast variety of hosts and systems, and such is the reason why CRISPR has gained so much attention from media and researchers alike. One type of application of CRISPR systems is in drug discovery, where molecular biology processes can be manipulated using CRISPR to aid in narrowing down potential therapeutic candidates (Shi et al., 2015). One such process is using isogenic knockouts to allow for swift establishment of causal roles for tumor suppressors, oncogenes, and other genetic factors in a particular disease of interest, demonstrated by Shi et al. through screening of protein domains. Homology-dependent repair (HDR) can induce knock-ins of mutant alleles to test the effects of disease-associated mutations in an isogenic environment, and CRISPR-Cas9 could be used to mediate gene knock-outs to facilitate testing of synthetic lethality in a drug candidate for tumor cells. Establishment of isogenic series also serves to analyze the effects that mutations have on the development of a certain disease, or to test the specificity of therapeutic candidates that target mutant genes (Richardson et al., 2016).

CRISPR in Public Health: Horizontal Gene Transfer

Antibiotic-resistant pathogens are a rampant problem in hospitals around the world, and the conferral of antibiotic resistance in pathogens that are normally vulnerable to antibiotics comes from horizontal gene transfer (HGT) (Marraffini and Sontheimer, 2008). HGT is the transfer of genes between two bacteria outside of the context of reproduction such as plasmid transformation, and is known to confer traits that are beneficial for survival under selective pressures such as immune responses of eukaryotic hosts or exposure to antibiotics (Gophna et al., 2015). CRISPR systems have been demonstrated as viable barriers to HGT, as CRISPR interference prevents the transformation of unencapsulated and avirulent pneumococci into strains that are capsulated and virulent during infection in mice (Bikard et al., 2012). Pneumococcal genomes do not harbor CRISPR loci, but Bikard et al. have demonstrated that introduction of CRISPR sequences from streptococci into *S. pneumoniae* results in prevention of DNA transformation of pneumococci. On the other hand, strong selective pressures for uptake of virulence and antibiotic resistance genes can result in loss of CRISPR protection in bacterial pathogens (Gophna et al., 2015). CRISPR loci and their DNA targets cannot coexist in the same host, as the introduction of CRISPR systems with spacers that target host chromosomal sequences results in the death of the host. In cases where the target DNA instills strong selective advantages such as antibiotic resistance, the selection survivors will have acquired the target DNA after loss of CRISPR interference activity or if there were mutations in the target DNA that would confer CRISPR interference evasion. Maintenance of genome integrity is a crucial role of

CRISPR systems, as HGT is a primary manner of obtaining new genetic material in prokaryotes (Bikard et al., 2012). On the other hand, CRISPR systems that target sequences with traits that are advantageous for selection may hinder the ability for the bacterial host to adapt and evolve under environmental selective pressures. Also, when CRISPR systems target sequences that confer adaptive traits, strains that possess deleted or inhibited CRISPR loci are selected for survival. However, these observations are only seen at the population level, as there is no evidence of HGT inhibition at the evolution level due to the mobility of CRISPR arrays, so that the presence or absence in extant genomes is not conclusive towards the long-term impact of CRISPR against horizontal gene transfer (Gophna et al., 2015). In addition, spacer acquisition is preferential towards new spacers from genomes that possess DNA sequences that match pre-existing spacers. This results in a strong bas towards sequences that are frequently encountered such as highly infectious viruses, as opposed to preference for the complete spectrum of exogenous DNA. Surveys of spacers in CRISPR arrays have shown that most of the matches are to prevalent archaeal viruses and bacteriophages, and numerous spacers arising from the same viral genome are typically present within the same CRISPR array.

Project Aims

Cmr Complex Mutants

Mutant constructs of Tma Cmr have been cloned with specific mutations to abolish RNase or DNase activity to test how each activity is modulated by the other. The RNase-dead mutant has a substitution in aspartate-26 for alanine in the Cmr4 subunit, and the DNase-dead mutant has a substitution in histidine-32 and aspartate-33 both to alanines in the HD nuclease domain of the Cmr2 subunit. Another mutant construct was made in the GGDD motif of Cmr2 where aspartate-585 and -586 are substituted for alanine to shed light on ambiguous published data on the motif. These mutants will be tested using activity assays to ascertain the effects that abolishment of RNase or DNase activity has on the overall activity of the Cmr complex.

Cryo-EM of Cmr Complex

The goal of obtaining a cryo-EM structure of the Cmr complex is to determine how binding to target DNA activates DNase activity, as the target binding site and the DNase active site are distant from each other in the complex. The structure could also help elucidate the mechanism of determining self from nonself targets, which is also a goal in the crystallization project.

Thermotoga maritima Csm1:4 and Cmr2:3 Crystallization

Thermophilic organisms are thought to have more thermostable proteins due to their ability to thrive in hotter and harsher climates. These thermostable proteins are advantageous to protein purification and crystallization as the proteins are less liable to precipitate or unfold at room temperature. All currently-published structures of the Cmr2:3 heterodimer are from *Pyrococcus furiosus* (Pfu), hence the determination of *Thermotoga maritima* (Tma) Cmr2:3 could yield information on how the heterodimer differs across organisms, as the amino acid sequence of Cmr2:3 is not well-conserved. When comparing sequences between Pfu and Tma, there are sequences that correspond to domains in the Pfu structure that do not exist in the Tma sequence, which is useful for mutagenic studies to determine the necessity and function of the domain(s) in question.

METHODS & MATERIALS

Expression of Cmr Complex Subunits

The expression of each Cmr subunit was identical to the protocol for Csm subunits, but Cmr6 was exclusively tagged in Cmr complex expression and purification as it will not properly express without the SUMO solubility tag. After overnight incubation of T7 Express cells transformed with plasmid containing one of the Cmr2 – 6 subunits, multiple colonies were picked from the lawn of T7 colonies and grown in a 10mL overnight starter culture with the appropriate antibiotic at 37°C. 1L of LB was inoculated for each Cmr2 to Cmr6 subunit with 5mL of overnight starter culture and grown at 37°C until the optical density (OD_{600}) of the culture reached 0.3. Upon reaching an OD_{600} of 0.3, protein expression was induced in the liter culture by addition of 200µL of 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown overnight at 20°C.

Purification of Cmr Complex

Each individual overnight liter culture was spun down at 2,500 RPM for 45 minutes at 4°C, followed by mixing of each cell pellet together (Estrella et al., 2016) and collective pellet resuspension in lysis buffer composed of 1M KCl, 20mM tris pH 8.0, 10mM imidazole, and 1mM TCEP. For each milliliter of resuspended T7 cells, 1µL of each protease inhibitor was added: 100mM PMSF, 1mM E-64, 2.5mM pepstatin A, and 1.7mM bestatin. The resuspended cells were lysed via microfluidizer followed by incubation of the crude cell lysate at 82°C for 10 minutes and then centrifugation of the incubated lysate at 18,000 RPM for 45 minutes at 20°C. After centrifugation, the

separated lysate supernatant was injected onto a charged IMAC column and the flow-through was discarded. The column was then washed with 5 column volumes of lysis buffer, repeated at least once more until protein elution was minimal from the wash steps. Bound protein was eluted with 10mL of elution buffer composed of 500mM KCl, 20mM tris pH 8.0, 250mM imidazole, and 1mM TCEP. The IMAC eluent was then injected onto a HiLoad™ 26/600 Superdex™ 200pg size exclusion column equilibrated in 350mM KCl, 20mM tris pH 8.0, and 1mM TCEP and collected fractions corresponding to the appropriate protein molecular weight.

Cmr1 Expression

After overnight incubation of T7 Express cells transformed with pHAT4-Cmr1, multiple colonies were picked from the lawn of T7 colonies and grown in a 10mL overnight starter culture with ampicillin at 37°C. 1L of LB was inoculated with 5mL of overnight starter culture and grown at 37°C until the OD_{600} of the culture reached 0.3. Upon reaching an OD_{600} of 0.3, protein expression was induced in the liter culture by addition of 200µL of 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown overnight at 20°C.

Cmr1 Purification

The overnight liter culture was spun down at 2,500 RPM for 45 minutes at 4°C, followed resuspension of the cell pellet in lysis buffer composed of 1M KCl, 20mM tris pH 8.0, 10mM imidazole, and 1mM TCEP. For each milliliter of resuspended T7 cells, 1µL of each protease inhibitor was added: 100mM PMSF, 1mM E-64, 2.5mM pepstatin

A, and 1.7mM bestatin. The resuspended cells were lysed via microfluidizer followed by incubation of the crude cell lysate at 82°C for 10 minutes and then centrifugation of the incubated lysate at 18,000 RPM for 45 minutes at 20°C. After centrifugation, the separated lysate supernatant was injected onto a charged IMAC column and the flow-through was discarded. The column was then washed with 5 column volumes of lysis buffer, repeated at least once more until protein elution was minimal from the wash steps. Bound protein was eluted with 10mL of elution buffer composed of 500mM KCl, 20mM tris pH 8.0, 250mM imidazole, and 1mM TCEP. The IMAC eluent was then injected onto a HiPrep™ 26/10 Desalting column via FPLC and flow-through was collected corresponding to the protein peak in the FPLC system chromatogram. 100µL of 75µM SENP was added to the desalt eluent and incubated at 37°C for at least 1 hour. The SENP digest was then injected onto a pre-charged IMAC column and collected the tag-cleaved Cmr1 from the flow-through. The IMAC flow-through containing the tag-cleaved Cmr1 was then injected onto a HiLoad™ 26/600 Superdex™ 200pg size exclusion column equilibrated in 350mM KCl, 20mM tris pH 8.0, and 1mM TCEP and collected fractions corresponding to the appropriate protein molecular weight.

Purification and Crystallization of Cmr2:3 Heterodimer

Separate cell pellets from 3L LB cultures of each codon-optimized pSATL-Cmr2 and codon-optimized pRSFd-Cmr3 were mixed and purified per the Cmr complex purification protocol, with additional steps after IMAC elution involving injection of the eluent onto a HiPrep™ 26/10 Desalting column via FPLC and collected flow-through corresponding to the protein peak in the FPLC system chromatogram. 100µL of 75µM

SENP was added to the desalt eluent and incubated at 37°C for at least 1 hour. The SENP digest was then injected onto a pre-charged IMAC column and collected the tag-cleaved protein from the flow-through, which was then injected onto the HiLoad™ 26/600 Superdex™ 200pg size exclusion column. Fractions specific to the Cmr2:3 molecular weight were pooled and concentrated. The concentrated protein sample was then used to set up 96-well Qiagen JCSG+ crystallization screens at a ratio of 200nL:200nL of protein:condition and incubated at 20°C.

Analytical Size Exclusion of Complete Cmr Complex

After pooling and concentrating size exclusion fractions corresponding to Cmr2-6, the purified complex was mixed with indicated crRNA and separately-purified Cmr1 in a >100μL reaction mixture. The final reaction condition contained 100mM KCl and 20mM tris pH 8.0, which was then incubated at 60°C for 10 minutes. The reaction mixture was injected onto an analytical Superose 6 Increase 10/300 GL column and collected fractions corresponding to the molecular weight of the full complex.

Ligation-Independent Cloning: Generation of Vector Inserts

When studying proteins from thermophilic organisms, there are cases where it is unfeasible or impossible to culture the organisms through current methods, so recombinant expression of the proteins in *E. coli* is used instead as a simpler and more cost-efficient method. One way of creating such recombinant plasmids is through ligation-independent cloning (LIC), which uses linearized vectors with amplified genes of interest that anneal through their generated sticky ends to form recombinant plasmids

that express our gene of interest (Figure 5). To express the target proteins in *E. coli*, the genes from the thermophilic organisms are cloned into expression vectors and transformed into the *E. coli* expression cells. To generate LIC-compatible ends, linearized vector and insert were mixed with T4 DNA Polymerase to exploit its 3'-5' exonuclease activity to create overhangs that are complementary to the vector or insert. A specific deoxynucleotide triphosphate (dNTP) was added to the reaction resulting in an equilibrium of 3'-5'-exonuclease and 5'-3'-polymerase activity to limit the exonucleolytic activity to the first complementary base. Our subunits of interest were either cloned into the pSATL vector with polyhistidine and small ubiquitin-like modifier (SUMO) tags, as well as with ampicillin antibiotic resistance genes encoded, or the pRSFL vector that encodes kanamycin antibiotic resistance. The 5' overhang created on the gene insert anneals with the 3' overhang of the vector which will result in an N-terminal His-SUMO tag on genes of interest expressed in pSATL. Both pSATL and pRSFL are T7-inducible vectors that contain compatible origins of replication and antibiotic resistance genes to allow for co-transformation and resultant double antibiotic selection of colonies.

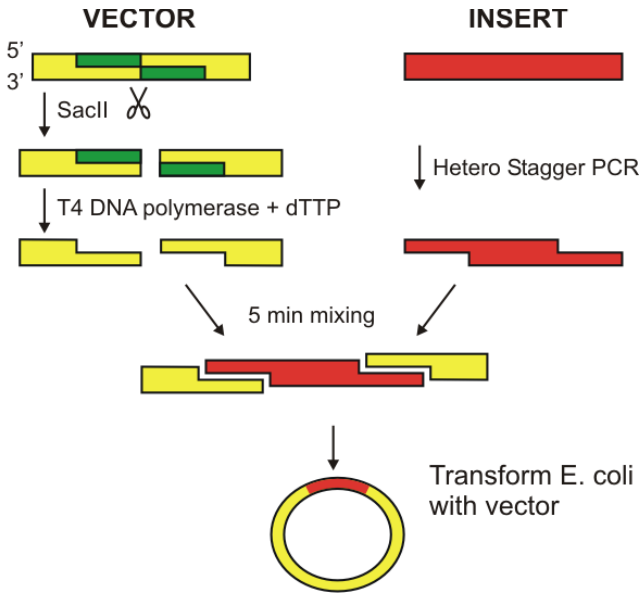


Figure 5: Schematic of generation of LIC-compatible ends and annealing of insert with vector (Utrecht University Department of NMR Spectroscopy, Bijvoet Center For Biomolecular Research).

Subunit Gene Amplification

To amplify genes encoding the Csm subunits from *Thermotoga maritima* genomic DNA, primer stocks (Figure 8 and 9) were diluted to 10μM and included in the reaction mixture detailed in Figure 6. The GC Reaction Buffer was selected due to the high GC content of the primers, as other reaction buffers used for PCR resulted in failure of gene amplification. The following PCR protocol was used to successfully amplify each subunit of the Csm interference complex from *Thermotoga maritima*: 1 pre-cycle at 98°C for 30 seconds, 35 denaturation cycles at 98°C for 10 seconds, anneal at 65°C for 20 seconds, and extend at 72°C for 15-30 seconds per kilobase pair, followed by 1 post-cycle at 72°C for 4 minutes, to which the reaction was then stored at 4°C until ready to collect. After electrophoretic separation of the PCR products, each individual band corresponding to the correct nucleotide length of each subunit was excised and followed manufacturer protocols of Thermo Scientific GeneJET Gel Extraction Kit for recovery of the insert.

PCR of Subunit Sequence	
Component	Volume (µL)
Deionized H_2O	33
6X GC Reaction Buffer (New England Biolabs)	10
10µM Forward Primer	2.5
10µM Reverse Primer	2.5
200µM dNTPs	1
Thermotoga maritima MSB8 Genomic DNA (ATCC® 43589D)	0.5
2.0 Units/µL Phusion Polymerase	0.5

LIC pRSFL Vector	
Component	Volume (µL)
Deionized H_2O	11.5
10X NEB 2.1 Buffer (New England Biolabs)	5
25mM dCTP	5
100mM DTT	2.5
Gel-purified pRSFL Vector	25
3.0 Units/µL T4 DNA Polymerase (New England Biolabs)	1

LIC Subunit Insert (pRSFL)	
Component	Volume (µL)
Deionized H_2O	9.6
10X NEB 2.1 Buffer (New England Biolabs)	2
25mM dGTP	2
100mM DTT	1
Gel-purified PCR Product	5
3.0 Units/µL T4 DNA Polymerase (New England Biolabs)	0.4

Figure 6: Reaction mixtures for the PCR amplification of subunit sequences from genomic DNA and generation of LIC-compatible ends in pRSFL vector and in the subunit insert

LIC pSATL Vector	
Component	Volume (µL)
Deionized H_2O	4.6
10X T4 DNA Polymerase Buffer (New England Biolabs)	2
25mM dTTP	2
100mM DTT	1
Gel-purified pSATL Vector	10
3.0 Units/µL T4 DNA Polymerase (New England Biolabs)	0.4

LIC Subunit Insert (pSATL)	
Component	Volume (µL)
Deionized H_2O	9.6
10X T4 DNA Polymerase Buffer (New England Biolabs)	2
25mM dATP	2
100mM DTT	1
Gel-purified PCR Product	5
3.0 Units/µL T4 DNA Polymerase (New England Biolabs)	0.4

Figure 7: Reaction mixture for generation of LIC-compatible ends in pSATL vector and in the subunit insert

TmCsm1.pSATL.F **CAACAGCAGACGGGAGGT**atgaaagacagagaagagcttg
TmCsm1.pSATL.R **GCGAGAACCAAGGAAAGGTTATTA**gctttttcttgagagaagatcc

TmCsm1.pRSFL.F **TTTAAGAAGGAGATATAGATC**atgaaagacagagaagagcttg
TmCsm1.pRSFL.R **TTATGGAGTTGGGATCTTATTA**gctttttcttgagagaagatcc

Figure 8: Sequences of primers used to PCR amplify Csm1 subunit sequence from *Thermotoga maritima* genomic DNA. Capitalized and bolded sections indicate the primer region that anneals to the vector, and lower case indicate the primer region that anneals to the Csm1 sequence.

Cloning a C-terminal His-tag into pRSFL-TmCsm5

5'-GTTGGAAGATTCCCAATCTACGGGATCCCACCATCACCATCACCATTAATAAGATCCCAACTCCATAAGTCTGGTAAAG-3'
5'-CAACCTTCTAAGGGTTAGATGCCCTAGGGTGGTAGTGGTAGTGGTAATTATTCTAGGGTTGAGGTATTCAGACCATTTC-5'

TmCsm5.cHis.F	CACCATCACCATTAATAAGATCCCAACTCCATAAG
TmCsm5.cHis.R	ATGGTGGGATCCCGTAGATTGGGAATCTTCCAAC

Figure 9: (Top) Sequence of pRSFL-Csm5 with C-terminal His-tag. Each region is color-coded with the green sequence being the C-terminus of Csm5, red sequence the added His-tag, and blue being the pRSFL vector. (Bottom) Sequences of primers used to PCR amplify Csm5 subunit sequence from *Thermotoga maritima* genomic DNA and clone into pRSFL vector with an added C-terminal His-tag. Regions of annealing are underlined in both the primers and the construct.

Generation of LIC-Compatible Ends

To generate LIC-compatible ends in pRSFL, the vector was first linearized with EcoRV and the product separated from uncleaved substrates by gel electrophoresis. The gel band corresponding to the length of pRSFL was excised and purified to be used in the reaction mixture detailed in Figure 6. The reaction for generating LIC-compatible ends of the PCR products was similar in formulation but differs in volumes and the deoxynucleotide used (Figure 6). Both LIC reactions were then incubated in a thermocycler at 22°C for 30 minutes followed by 75°C for 20 minutes to produce LIC-compatible ends of the vector and insert. The LICed vector and inserts were then combined at 2µL of LICed PCR insert with 2µL of LICed vector with 6µL of water and incubated at room temperature for 5 minutes, and a no-insert reaction with water instead of insert was prepared as a negative control. Top10 Chemically Competent *E. coli* cells (Invitrogen) were then transformed with the annealed construct, and a no-insert control was included as well. Transformation of Top10 cells consisted of adding 10µL of annealed LIC mixture to the competent cells on ice, incubation on ice for 10 minutes, heat shock at 42°C for 30 seconds, incubation on ice for 10 minutes again, addition of 1mL of Luria-Bertani (LB) broth to cells and incubation at 37°C for 20 minutes, and then spreading of incubated cells on kanamycin resistance LB agar plates and grown overnight at 37°C.

The LIC protocol for the pSATL vector was procedurally identical to the pRSFL protocol with minor differences such as deoxynucleotide choice in the generation of single-stranded overhangs in the gene insert and the vector. First, StuI restriction enzyme was used instead of EcoRV for linearization of the pSATL vector, and ampicillin

resistance LB plates were used to grow the transformed Top10 cells instead of kanamycin. The LIC reaction mixtures were also different for the vector and insert, as detailed in Figure 7.

To determine whether the isolated colonies growing on the resistance plates possess the plasmid of interest, a colony screen was performed by resuspension of 1 picked colony in 100μL of autoclaved water, running the genomic PCR reaction but replacing genomic DNA stock with 0.5μL of resuspended colony, and then running the PCR product on an agarose gel to confirm that the band corresponds to the correct kilobase length.

Expression of Csm Subunits

After transforming Top10 cells with LIC-cloned plasmids containing each Csm subunit, an isolated colony of Top10 cells was picked and grown in a 5mL LB culture with ampicillin overnight at 37°C. Using a standard minipreparation kit, plasmid was extracted and purified from the 5mL overnight culture, and this purified plasmid was used to transform T7 Express Competent *E. coli* cells (New England Biolabs) for protein expression as opposed to the plasmid maintenance Top10 strain. After overnight incubation, multiple colonies from the lawn of T7 colonies were taken and inoculated into a 10mL overnight starter culture with ampicillin at 37°C. 1L of Terrific Broth (TB) was then inoculated with 5-10mL of overnight starter culture and grown at 37°C until the optical density (OD_{600}) of the culture reached 0.8. Upon reaching an OD_{600} of 0.8, protein expression was induced in the liter culture by addition of 1mL of 1M isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown overnight at 20°C.

Purification of Csm Subunits

The overnight liter cultures were spun down at 2,500 RPM for 45 minutes at 4°C and the cell pellet was resuspended in lysis buffer composed of 350mM NaCl, 20mM tris pH 8.0, 10mM imidazole, and 10% v/v glycerol. For each milliliter of resuspended T7 cells, 1µL of each protease inhibitor was added: 100mM phenylmethylsulfonyl fluoride (PMSF), 1mM proteinase inhibitor E-64, 2.5mM pepstatin A, and 1.7mM bestatin. The resuspended cells were lysed via microfluidizer followed by incubation of the crude cell lysate in an 80°C water bath for 20 minutes until lysate temperature reaches 71°C, followed by centrifugation of the incubated lysate at 18,000 RPM for 45 minutes at 4°C. To charge immobilized metal affinity chromatography (IMAC) resin for affinity chromatography, 100mM nickel acetate was applied to the column and the column was washed with at least 2 column volumes of water, followed by 5 column volumes of lysis buffer to equilibrate the column. After centrifugation, the separated lysate supernatant was injected onto the IMAC column and the flow-through was discarded. The column was then washed with 5 column volumes of lysis buffer to remove unbound proteins, and was repeated at least once more until protein elution was minimal from the wash steps. Bound protein was then eluted with 10mL of elution buffer composed of 350mM NaCl, 20mM tris pH 8.0, 250mM imidazole, and 10% v/v glycerol, followed by injection of the eluent onto a HiPrep™ 26/10 Desalting column via fast protein liquid chromatography (FPLC) and flow-through was collected corresponding to the protein peak in the FPLC system chromatogram. 100µL of 75µM SUMO-specific protease (SENP) was added to the desalt eluent and incubated at 37°C for at least 1 hour to cleave His tags from the respective protein substrate. The SENP digest was then injected onto a pre-charged

IMAC column and the tag-cleaved protein collected from the flow-through, and any bound tagged proteins as well as bound SENP from the IMAC resin was eluted by flushing the column with elution buffer. The flow-through was injected onto a HiLoad™ 26/600 Superdex™ 200pg size exclusion column equilibrated in 350mM KCl, 20mM HEPES pH 7.5, and 1mM tris(2-carboxyethyl)phosphine (TCEP) and certain eluted fractions were collected corresponding to the appropriate protein molecular weight.

Purification and Crystallization of Csm1:4 Heterodimer

T7 Express cells were co-transformed with pSATL-Csm1 and pRSFL-Csm4 and grown in 6L of ampicillin/kanamycin double antibiotic Terrific Broth (TB) cultures. Following the Csm complex purification protocol and concentration of final size exclusion fractions, 25µL of 20mg/mL purified protein was produced to set up a 96-well Qiagen JCSG+ crystallization screen at a ratio of 200nL:200nL of protein to crystallization solution in each well. The crystallization tray was then incubated at 20°C.

RESULTS

Cmr Complex Purification

Purification of the Cmr complex was successful based on previously-formulated expression and purification protocol by Michael Estrella (Estrella et al., 2016), but has yielded inconsistent expression of Cmr2. We then hypothesized that codon-optimizing the Cmr2 sequence for expression in *E. coli* could potentially alleviate the inconsistent expression issue. Codon optimization played a role in the recombinant expression of our proteins, as most amino acids were encoded by more than one codon triplet some organisms will use one codon more often than another. Purification of the Cmr complex involved combining cell pellets expressing each subunit, lysis by microfluidizer, incubation at 80°C to facilitate complex formation, followed by IMAC affinity chromatography and size exclusion to yield ~180µg of purified protein from 5L of LB (Figure 10). The consistent expression of codon-optimized Cmr2 validated our hypothesis that codon optimization would fix inconsistent expression issues in Cmr2. These mutant complexes will be used in biochemical activity assays and the D26A construct will also be used for cryo-EM to obtain structures of the Cmr complex bound to different RNA targets.

To better understand how binding of RNA targets influence the overall conformation of the Cmr interference complex, RNase-deficient constructs were made. Aspartate-26 of Cmr4 was mutated to alanine in the D26A mutant, Cmr2 histidine-32 to alanine and aspartate-33 to alanine mutations were made in the HD mutant construct, and Cmr2 aspartate-585 as well as aspartate-586 were mutated to alanine in the GGDD palm

mutant. The D26A mutant was deficient in RNase activity as the catalytic residue was mutated, and allowed for binding of the RNA transcript without cleavage occurring (Ramia et al., 2014). The GGDD mutant aimed to shed light on the unknown role of the palm domain of Cmr2. Wild-type Cmr complex was also purified as a control.

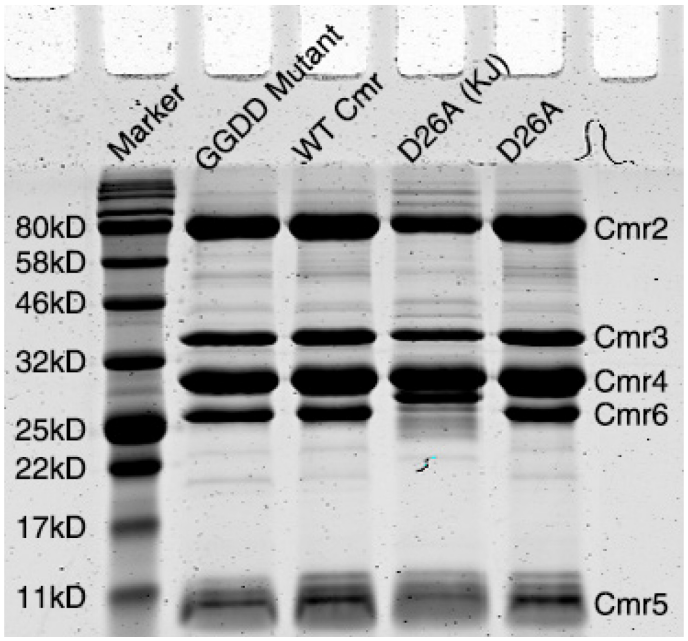


Figure 10: 15% SDS-PAGE gel of 2.5μM samples of Cmr2 – 6 constructs. D26A (KJ) prepared by Kaitlin Johnson in the Bailey lab with a Cmr6 expressed in a pHAT4 vector encoded a longer N-terminal His tag, compared to the His tag from pHAT2 which was used to express Cmr6 in the other constructs. This explained why the Cmr6 band was at a higher molecular weight compared to the other three constructs

Cryo-EM of Cmr Complex

The goal of obtaining a cryo-EM structure of the Cmr complex was to determine how binding to target RNA activates DNase activity, as the target RNA binding site and the DNase active site are distant from each other in the complex. The structure could also help elucidate the mechanism of determining self from nonself targets by obtaining structures bound to self and nonself RNA targets. Existing crystal structures of hybrid *Pyrococcus furiosus*/*Archaeoglobus fulgidus* Cmr complex (Osawa et al., 2015) and cryo-EM structures of *Thermus thermophilus* Cmr complex (Taylor et al., 2015) do not contain the HD domain of Cmr2, and thus information on inhibition/activation of DNase activity cannot be deduced from existing structures of the Cmr complex. The D26A mutant of the Cmr complex was of importance as it allowed for binding of RNA without cleavage of the RNA target, as DNase activity would be turned off if the RNA target were to be cleaved (Estrella et al., 2016). The RNA-bound structure could then be compared with the unbound structure to determine any conformational changes that would occur after binding of the RNA target.

Cmr complex with D26A mutation was purified and mixed with crRNA and Cmr1 in a 1:4:15 ratio of D26A:crRNA:Cmr1. Thermophiles like *Thermotoga maritima* grow at 80°C and exhibit the most biochemical activity at that temperature, which was the target temperature to incubate Cmr1 – 6 with crRNA. The mixture was incubated at 60°C for 10 minutes to facilitate complex formation, as an 80°C incubation temperature resulted in precipitation of the mixture. With the goal of chromatographically separating the full Cmr1-6 with crRNA complex from all other mixture components, the incubated mixture was injected onto an analytical Superdex™ 200 Increase 10/300 GL column in a

buffer of 100mM KCl and 20mM tris pH 8.0, and buffers with higher salt concentrations resulted in weaker association of Cmr1 to the Cmr2 – 6 complex. Analytical size exclusion results yielded peaks corresponding to the full complex at ~10.5mL (Figure 4), which coeluted with a mixture of complexes containing varied stoichiometries as seen in the peaks between 8 and 10mL elution volume. Cmr1 presence decreased sooner than the other subunits across the elution volume of the full complex when the size exclusion buffer had a higher salt concentration (Figures 11, 12, and 13), indicating that Cmr1 association with the complex was sensitive to salt concentration. Improved separation of the peaks was achieved through analytical size exclusion with a Superose 6 Increase 10/300 GL column (Figures 14 and 15), where the peak corresponding to the full Cmr complex had no observable overlaps with adjacent peaks, a positive sign towards monodispersity of the samples from the peak.

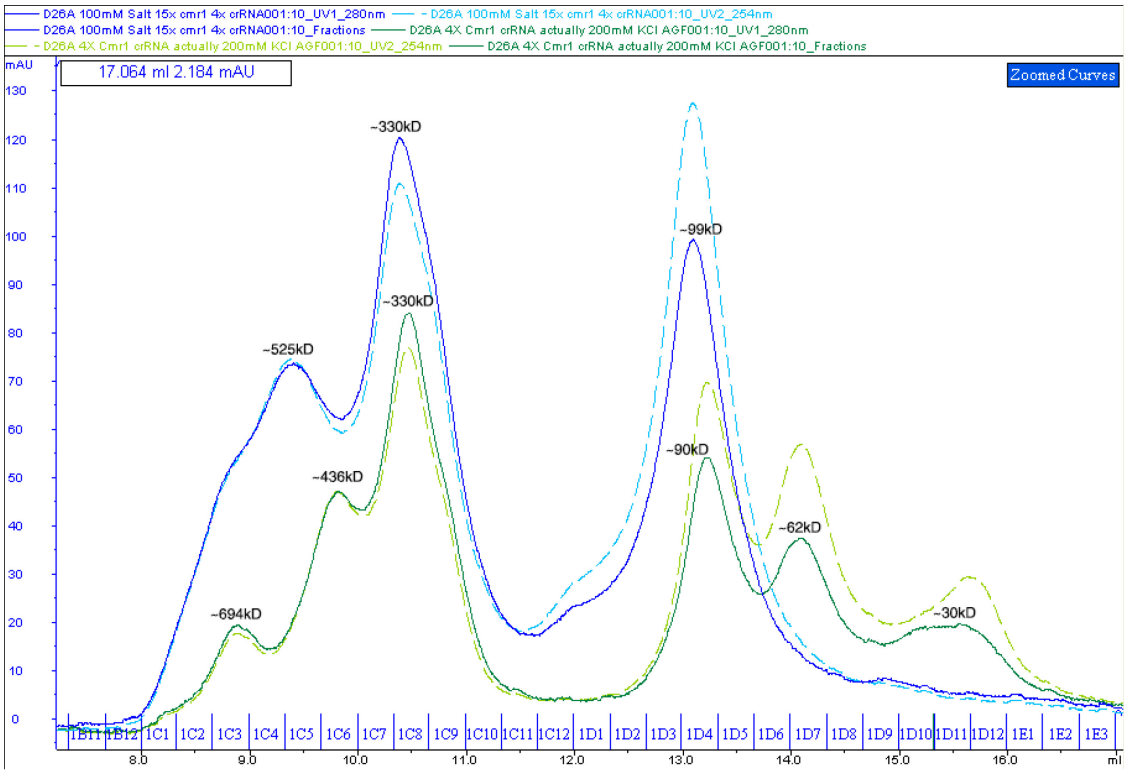
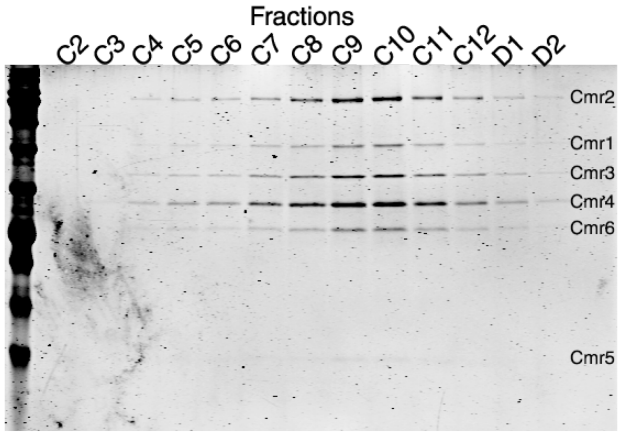
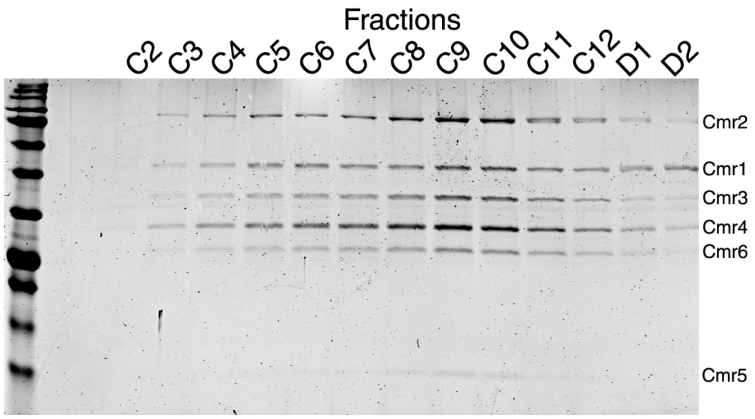


Figure 11: Analytical size exclusion chromatogram comparison of D26A complex at 1:15:4 ratio of Cmr2-6:Cmr1:crRNA complexed at 200mM KCl (green trace), and complexed at 100mM KCl (blue trace)



Fraction	Subunit	Raw Area	Area (Normalized to MW)	Percent of Total Lane Intensity	Normalized Intensity (to Cmr3)
C5	Cmr1	104.607	2.08	6.25%	0.42
	Cmr3	187.607	4.95	14.85%	1.00
C6	Cmr1	124.607	2.48	7.32%	0.43
	Cmr3	217.192	5.73	16.90%	1.00
C7	Cmr1	206.314	4.11	5.85%	0.32
	Cmr3	492.314	12.99	18.48%	1.00
C8	Cmr1	442.385	8.81	6.88%	0.38
	Cmr3	875.021	23.09	18.02%	1.00
C9	Cmr1	745.092	14.84	6.54%	0.36
	Cmr3	1568.435	41.38	18.23%	1.00
C10	Cmr1	845.799	16.85	7.53%	0.40
	Cmr3	1611.021	42.51	19.01%	1.00
C11	Cmr1	390.971	7.79	6.04%	0.36
	Cmr3	808.899	21.34	16.55%	1.00
C12	Cmr1	154.899	3.09	5.81%	0.31
	Cmr3	371.728	9.81	18.48%	1.00
D1	Cmr3	202.607	4.04	14.84%	1.00

Figure 12: (Top) 15% SDS-PAGE gel of analytical size exclusion fractions of Cmr D26A full complex formed at 200mM KCl. (Bottom) Quantification of Cmr1 band intensity from SDS-PAGE gel of analytical size exclusion fractions



Fraction	Subunit	Raw Area	Area (Normalized to MW)	Percent of Total Lane Intensity	Normalized Intensity (to Cmr3)
C3	Cmr1	200.728	4.00	13.68%	1.03
	Cmr3	146.607	3.87	13.23%	1.00
C4	Cmr1	455.263	9.07	14.77%	0.98
	Cmr3	351.728	9.28	15.11%	1.00
C5	Cmr1	927.678	18.48	16.70%	1.09
	Cmr3	644.749	17.01	15.38%	1.00
C6	Cmr1	1056.263	21.04	17.08%	1.15
	Cmr3	691.556	18.25	14.81%	1.00
C7	Cmr1	951.263	18.95	13.42%	0.86
	Cmr3	831.971	21.95	15.55%	1.00
C8	Cmr1	1027.385	20.47	11.03%	0.67
	Cmr3	1151.092	30.37	16.37%	1.00
C9	Cmr1	1574.971	31.37	11.45%	0.66
	Cmr3	1792.971	47.31	17.26%	1.00
C10	Cmr1	1548.678	30.85	12.81%	0.74
	Cmr3	1577.849	41.63	17.29%	1.00
C11	Cmr1	905.142	18.03	12.74%	0.75
	Cmr3	913.849	24.11	17.03%	1.00
C12	Cmr1	748.142	14.90	17.22%	1.04
	Cmr3	541.899	14.30	16.52%	1.00
D1	Cmr1	739.556	14.73	29.69%	2.21
	Cmr3	252.142	6.65	13.41%	1.00

Figure 13: (Top) 15% SDS-PAGE gel of analytical size exclusion fractions of Cmr D26A full complex formed in buffer containing 100mM KCl. (Bottom) Quantification of Cmr1 band intensity from SDS-PAGE gel of analytical size exclusion fractions

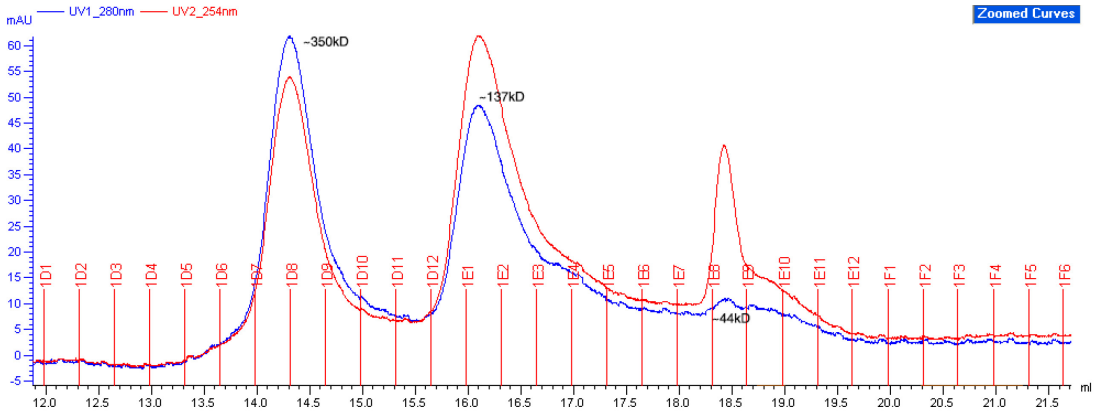
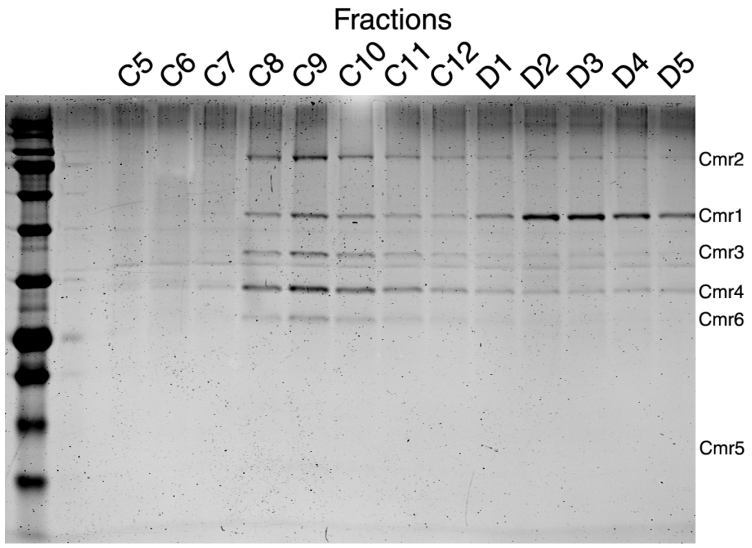


Figure 14: Superose analytical size exclusion chromatogram of Cmr D26A full complex formed in buffer containing 100mM KCl



Fraction	Subunit	Raw Area	Area (Normalized to MW)	Percent of Total Lane Intensity	Normalized Intensity (to Cmr3)
D8	Cmr1	521.385	10.39	9.72%	0.81
	Cmr3	488.435	12.89	12.07%	1.00
D9	Cmr1	1116.92	22.25	12.00%	0.89
	Cmr3	951.849	25.11	13.54%	1.00
D10	Cmr1	544.849	10.85	9.87%	0.63
	Cmr3	656.385	17.32	15.75%	1.00
D11	Cmr1	352.678	7.03	11.51%	0.90
	Cmr3	295.435	7.80	12.77%	1.00
D12	Cmr1	263.971	5.26	12.15%	1.19
	Cmr3	168.021	4.43	10.24%	1.00
E1	Cmr1	681.092	13.57	35.42%	4.77
	Cmr3	107.899	2.85	7.43%	1.00
E2	Cmr1	2331.627	46.45	67.29%	7.86
	Cmr3	224.092	5.91	8.57%	1.00
E3	Cmr1	2331.042	46.44	73.83%	9.69
	Cmr3	181.556	4.79	7.62%	1.00
E4	Cmr1	1657.335	33.01	65.07%	8.08
	Cmr3	154.849	4.09	8.05%	1.00
E5	Cmr1	1075.335	21.42	57.50%	18.54
	Cmr3	43.778	1.16	3.10%	1.00

Figure 15: (Top) 15% SDS-PAGE of Superose analytical size exclusion fractions from Cmr D26A full complex formed in buffer containing 100mM KCl. (Bottom) Quantification of Cmr1 band intensity from SDS-PAGE gel of analytical size exclusion fractions

Cmr2:3 Crystallization

Existing structures of Cmr2:3 from *Pyrococcus furiosus* do not contain the HD nuclease domain, and since the amino acid sequences of Cmr2 and Cmr3 are not well-conserved across organisms, there was merit to crystallizing *Thermotoga maritima* Cmr2:3 to obtain information that was not available in existing structures. The D2 domain in *Pyrococcus* Cmr2 makes contacts with the RNA target base-paired with the 5' end of the crRNA (Figure 16). The positioning of the D2 domain could also suggest involvement in binding to the 3' flanking sequence of the RNA target, which is the region of the target involved in self vs non-self recognition (Marraffini and Sontheimer, 2010). Sequence alignment of *Pyrococcus* and *Thermotoga* Cmr2 indicated that the D2 domain does not appear in the *Thermotoga* variant, which could suggest that there could be a different mechanism of distinguishing self vs non-self targets in *Thermotoga maritima* Cmr2. Mutagenic studies on D2 domain excision in *Pyrococcus* Cmr2 or D2 domain insertion in *Thermotoga* Cmr2 could potentially shed light on the functional role that the D2 domain plays in each organism. Codon-optimized Cmr2 expressed in pSATL and codon-optimized Cmr3 expressed in pRSFd are individually grown in 3L of LB each, where the centrifuged cell pellets were mixed, lysed, and purified together according to the Cmr complex protocol to form the stable Cmr2:3 heterodimer. Crystallization conditions from JCSG+ crystal screen yielded crystal florets from 20.35mg/mL (260/280 = 0.56) protein sample in a 200nL:200nL ratio with crystallization solution: 0.04M magnesium acetate, 0.05M sodium cacodylate pH 6.0, and 30% v/v MPD. Florets of crystal needles formed within 3 days of crystal drop setup (Figure 17), and the largest needles formed at 0.04M magnesium acetate, 0.05M sodium cacodylate pH 6.2, and 28%

v/v MPD. Numerous crystals were sent to the Stanford Synchrotron Radiation Lightsource (SSRL) for data collection. The crystals diffracted x-rays to 3.0Å resolution and belong to space group $P2_12_12_1$. Unfortunately, molecular replacement (performed by Scott Bailey) revealed that the crystals were not formed by the Cmr2:Cmr3 complex but by the Cmr3 subunit alone. We hypothesized that a component of the crystallization solution could have interfered with the heterodimer interface, or that the His-SUMO tag could have potentially disrupted the crystal packing mode of Cmr2:3. The heterodimer was then purified and the His-SUMO tag was cleaved before Qiagen JCSG+ and Nucleix crystallization trays were set up, and initial crystallization screens yielded ~200μm needle-like crystal growth in numerous conditions. The crystals that formed from the conditions identical to those that formed the Cmr3 crystal florets were ruled out, as they were likely to be only Cmr3 crystals. The other conditions that had crystal growth have not been followed up with secondary trays yet.

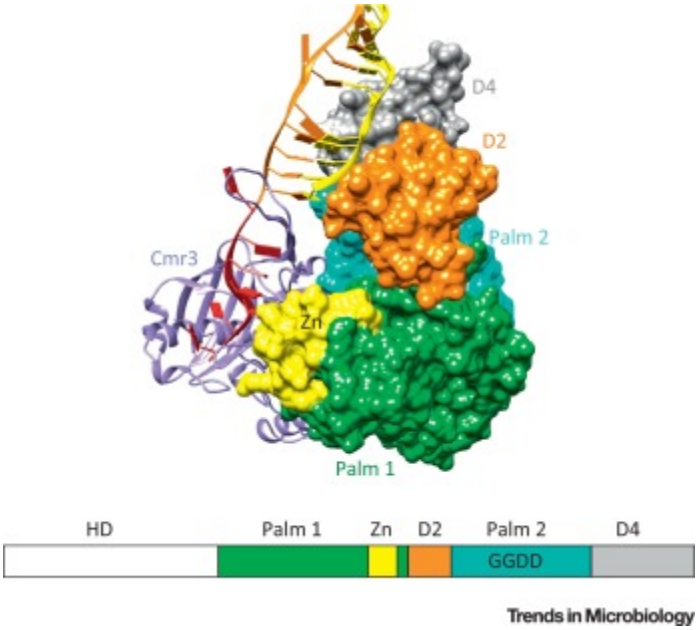


Figure 16: *Pyrococcus furiosus* Cmr2 with D2 domain in orange, RNA target in yellow, crRNA in red/orange, and Cmr3 subunit in purple (Review, Tamulaitis et al., 2016)

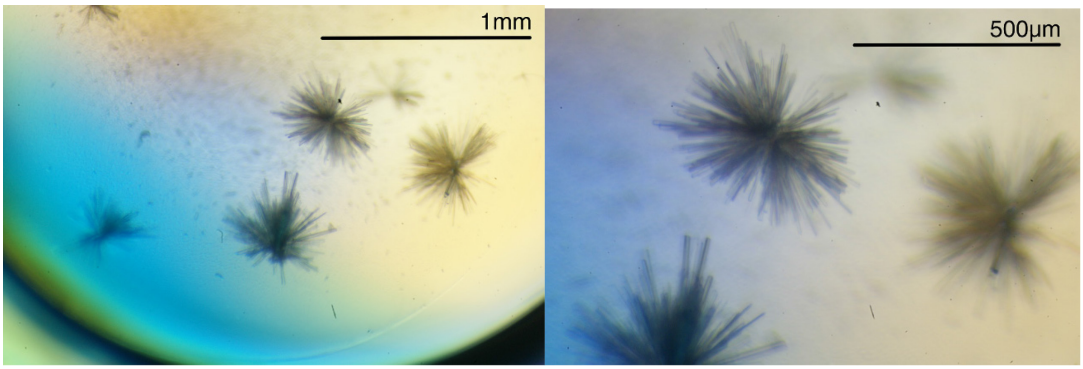


Figure 17: (Left) crystallization tray sitting drop with crystal florets of Cmr3 at 77x magnification. (Right) crystallization tray sitting drop with crystal florets of Cmr3 at 155x magnification

Csm Subunit LIC Cloning

When studying proteins from thermophilic organisms, there were cases where it was unfeasible or impossible to culture the organisms through current methods, so recombinant expression of the proteins in *E. coli* was used instead. To express the target proteins in *E. coli*, the genes from the thermophilic organisms were cloned into expression vectors and transformed into the *E. coli* expression cells. To clone each subunit of the Csm1 – 5 complex from *Thermotoga maritima* genomic DNA, we designed primers to anneal to the 15-20 nucleotides of the 5' and 3' ends of each Csm subunit and utilized polymerase chain reaction (PCR) to amplify the subunit from *Thermotoga maritima* DNA used as the PCR reaction template for the designed primers to anneal. We electrophoretically separated the PCR reaction products (Figure 18) and confirmed subunit length by its mobility on the agarose gel. Following ligation-independent cloning, each Csm subunit was successfully incorporated into both pSATL and pRSFL vectors and confirmed through dideoxy Sanger sequencing. Top10 Chemically Competent *E. coli* cells (Invitrogen) were chosen for transformation and cloning as they have recA activity silenced to reduce nonspecific recombination of cloned DNA, and have endA1 mutations that eliminated nonspecific cleavage by host endonuclease 1. These two features of the Top10 cell line helped facilitate consistent amplification and maintenance of recombinant plasmids.

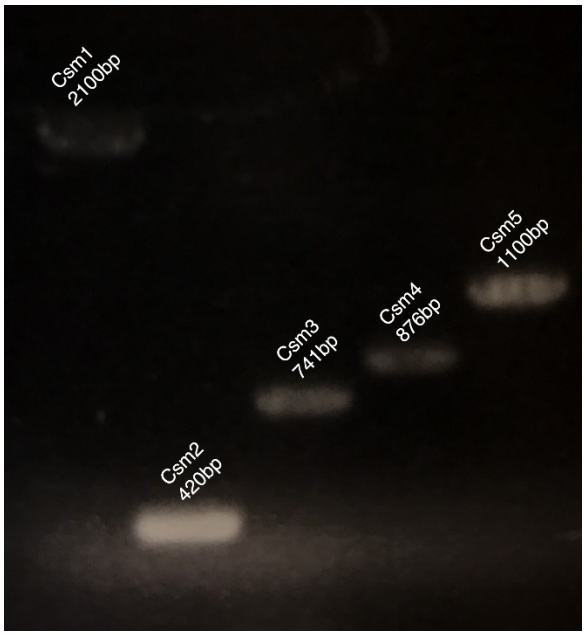


Figure 18: 1% agarose gel of PCR-amplified gene inserts

Csm Subunit Expression and Purification

Aside from individual subunits from other organisms, there was currently no published crystal structure of the Csm complex let alone from *Thermotoga maritima*, and much could be gained from analysis of the overall structure and how it compares to other CRISPR interference complexes. Since the expression and purification of *Thermotoga maritima* Csm subunits have not been previously performed in the lab, most of the initial protocols were adapted from published Cmr purification protocols (Estrella et al., 2016). With the goal of producing enough Csm complex to set up crystallization screens and perform biochemical assays, we separately expressed Csm1-5 subunits in 1L of Luria-Bertani (LB) media and mixed the centrifuged cell pellets together, which was then lysed by microfluidizer. Very little Csm complex formation was observed with current purification protocol.

Csm5 constructs were also designed with N-terminal and C-terminal polyhistidine (His) tags and LIC cloned into a pRSFL vector. This was done to determine whether the small ubiquitin-like modifier protein (SUMO) tag from pSATL was the cause for low protein concentrations after heat treatment of the crude cell lysate. We found that C-terminally His-tagged Csm5 expressed better than N-terminal His-SUMO and N-terminally His-tagged Csm5 constructs, which validated one of our hypotheses that the location of the His tag was important for expression of this subunit.

SDS-PAGE of the immobilized metal affinity chromatography (IMAC) elution showed enriched presence of only the His-tagged subunit, with faint or nonexistent bands corresponding to untagged subunits. Purification of each Csm subunit in 6L of media yielded undetectable levels of untagged subunits, but the C-terminally His-tagged Csm5

peak isolated from size exclusion yielded 420µg of protein which was a positive indication that C-terminal tagging of subunits may improve protein yield after purification (Figures 19 and 20).

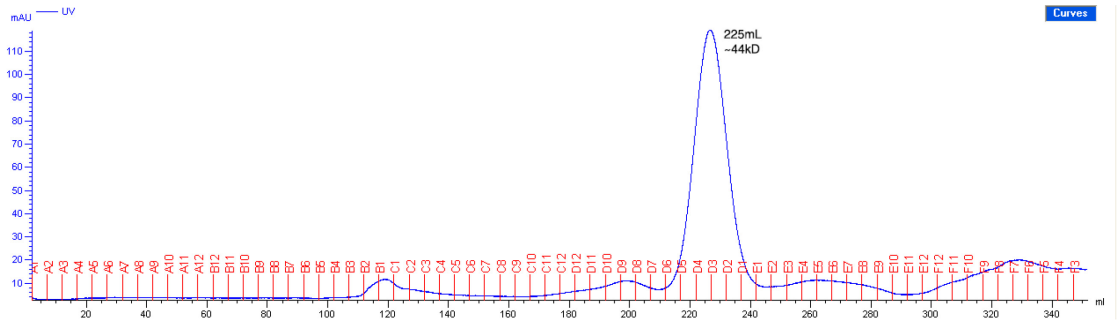


Figure 19: Csm complex size exclusion chromatogram, with Csm5-CHis protein peak corresponding to fractions D6 to E1.

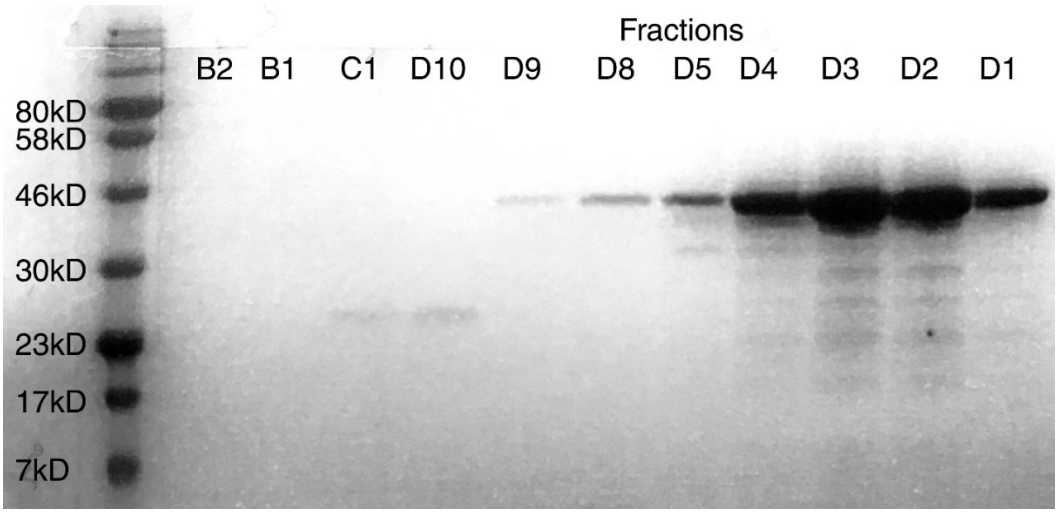


Figure 20: 15% SDS-PAGE gel of Csm5-CHis size exclusion fractions

Csm1:4 Purification

Since Csm1 and Csm4 are associated together in the Csm complex, we postulated that they would associate when co-expressed in *E. coli*. Once the complex has expressed, we hypothesized that the complex would co-elute from IMAC affinity columns together as a heterodimer if one subunit was His-tagged. When Csm1 and Csm4 were separately expressed and then purified together, minimal heterodimer formed but expression improved in the purification of double-transformed T7 Express cells with Csm1 and Csm4 grown in LB (Figures 21 – 23). Co-transformation of the heterodimer was performed in other expression cells such as ArcticExpress and Rosetta™ competent cells, but they did not yield any appreciable improvements in expression of the subunits. Other medias that were richer in growth nutrients were also tested, with 2X-YT media and Terrific Broth (TB) producing higher yields of protein, with a preference for TB media as it yielded the highest amount of protein produced across the three media types tested. Purification of the subunits expressed by T7 cells in TB media yielded 400µg of pure protein with a 260/280 absorbance ratio of 0.68 which corresponded to a sample with very little nucleic acid content.

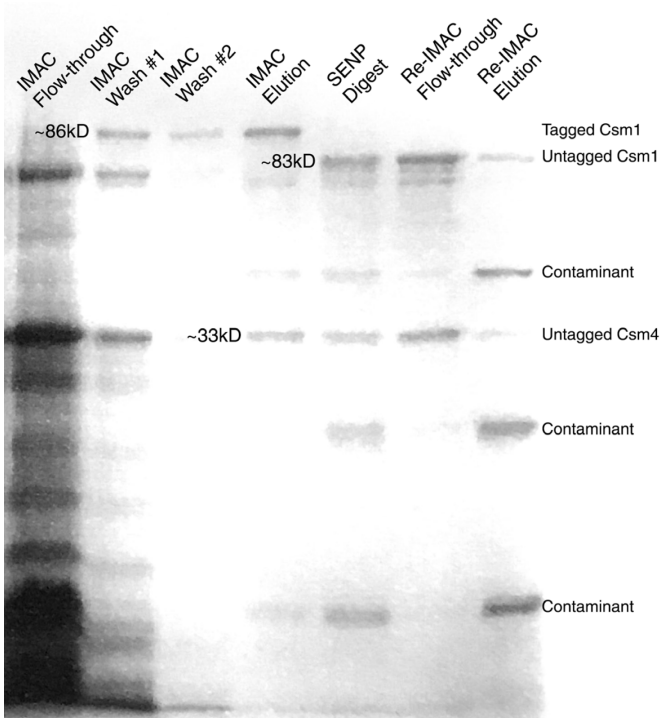


Figure 21: 15% SDS-PAGE gel of Csm1:4 IMAC purification and SENP digestion

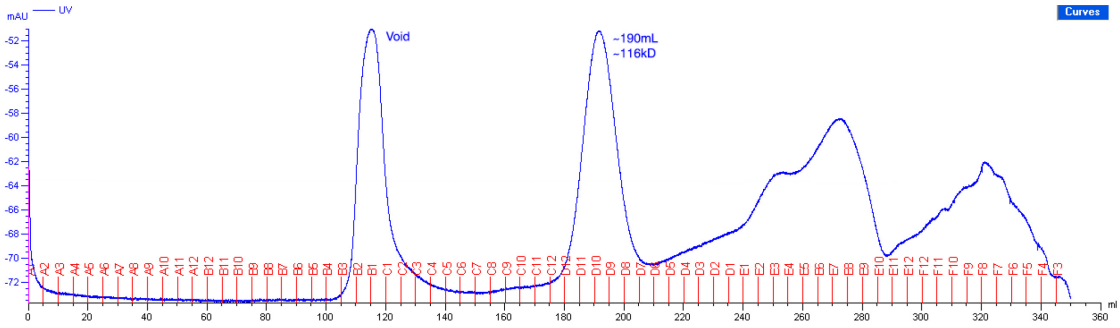


Figure 22: Size exclusion chromatogram of tag-cleaved Csm1:4

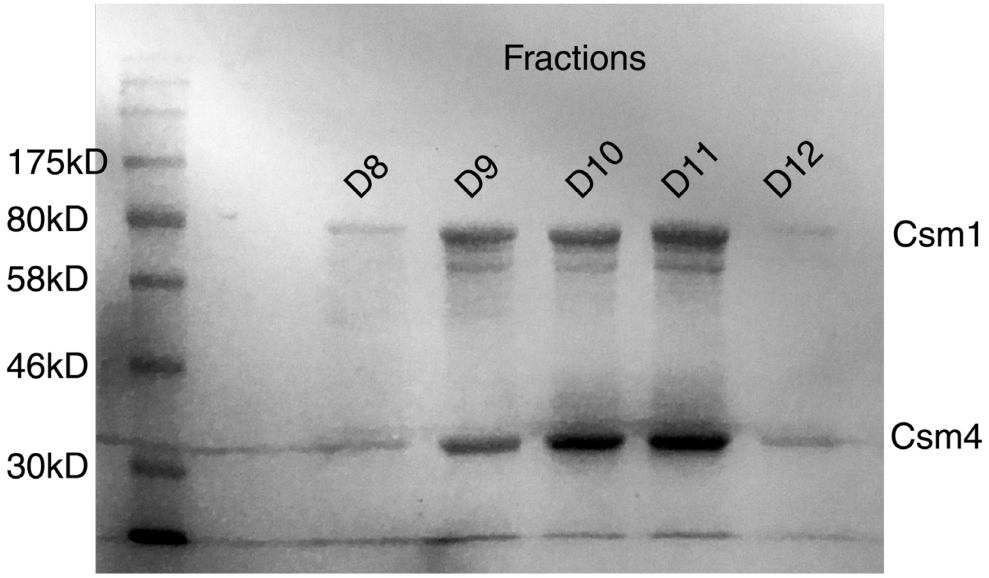


Figure 23: 15% SDS-PAGE gel of Csm1:4 size exclusion fractions D8 – D12. 250µL aliquots of each fraction were acetone precipitated and reconstituted in water

Csm1:4 Crystallization

The *Thermotoga maritima* Csm1:4 heterodimer was a prime target for crystallization and structural determination as there was no existing published structure of the complex, and would shed light on any structural homologies with other orthologs of the Cas10 – Cas5 complex. Upon expression and purification of Csm1:4 in 6L of TB media, 20μL of 20.48mg/mL protein was used to set up a 96-condition crystallization tray with Qiagen JCSG+ crystal condition screen at a ratio of 200nL:200nL protein:condition. Crystals formed at various conditions, with most continual growth in 0.2M Li_2SO_4, 0.1M sodium acetate pH 4.5, and 30% v/v PEG8000, as well as in 0.17M ammonium sulfate, 25.5% v/v PEG4000, and 15% v/v glycerol.

DISCUSSION

Cmr Complex Production

The purification protocol of the Cmr complex had been previously formulated and published by Michael Estrella, but past purifications of the Cmr subunits resulted in inconsistent expression of Cmr2. We tested whether codon-optimizing the Cmr2 sequence would alleviate the inconsistent expression issues, and concluded that codon optimization successfully alleviated the inconsistent expression issues. With the subunits already cloned into appropriate vectors, there were little issues and obstacles encountered during the expression and purification of Cmr complex mutants.

Cryo-EM of Cmr Complex Mutants

With the goal of producing purified monodisperse Cmr complex (Cmr2-6 with Cmr1) with bound crRNA, a reaction mixture of each component is incubated at 60°C and injected onto a Superose 6 Increase 10/300 GL column to separate the full complex from complexes with incomplete subunit stoichiometries. An initial obstacle was getting Cmr1 to associate with the Cmr2-6 complex, as SDS-PAGE of analytical size exclusion fractions across the peak corresponding to the full complex showed that band intensity corresponding to Cmr1 seemed to disappear at earlier fractions compared to the other Cmr2-6 subunits. Reaction mixtures formed in >200mM KCl resulted in suboptimal Cmr1 association with the rest of the complex, but reaction mixtures with a final KCl concentration of 100mM resulted in markedly reduced disappearance of Cmr1 at earlier fractions, so it can be deduced that Cmr1 association with the Cmr2-6 complex is sensitive to salt concentrations with an affinity for lower salt concentrations.

Quantification of band intensities from SDS-PAGE of analytical size exclusion fractions is not an accurate method of determination of subunit dissociation, as there are numerous sources of error in data collection and analysis such as the qualitative nature of integrating band intensities from gel scans. Due to the low molecular weight of Cmr5, there has been recurring issues with the subunit binding Coomassie Brilliant Blue dye in a sub-stoichiometric manner, so quantification of Cmr5 bands stained with Coomassie is inaccurate in representation of true Cmr5 stoichiometry in the complex. Monodispersity is also hard to gauge through SDS-PAGE analysis of the fractions, as it shows the total intensity of each subunit in the fraction instead of the intensity of each subunit associated in the full crRNA-bound Cmr complex within the fraction.

A common issue in sample preparation for cryo-EM is purifying enough protein to yield high concentrations of sample in a 3-5µL volume (Thompson et al., 2016), and since the process of producing monodisperse crRNA-bound Cmr complex involves numerous purification steps that inevitably reduce the final yield of purified sample, there are optimizations that potentially need to be made with the simplest solution being purification of a larger volume of cell cultures. Another issue with samples in cryo-EM is degradation of structural integrity of the Cmr complex, as a high vacuum is needed to prevent scattering of electrons by air molecules which results in disruption of aqueous protein sample preservation (Bai et al., 2014), a potentially serious issue that could increase the relative salt concentration of the Cmr complex sample and result in dissociation of the salt-sensitive Cmr1 subunit.

Another factor to account for is the stability of the crRNA, as the analytical size exclusion chromatograms indicate some degree of nucleic acid degradation in the

reaction mixture. Future efforts will require confirmation that the crRNA in the incubated complex is intact, as there are risks of incorporating truncated crRNA to the Cmr interference complex.

Future efforts involve activity assays to determine whether the full complex in the peak contains the correct number of subunits, and to test whether the crRNA within the full complex is intact.

Cmr2:3 Crystallization

The crystallization conditions that facilitated crystal floret formation of purified His-tagged Cmr2:3 was reproducible within an appreciable range of buffering pHs and percentages of MPD. Ultimately the crystals formed contained a crystal packing mode that could only accommodate Cmr3, and molecular replacement of the diffraction pattern confirmed that the crystal florets only contained Cmr3. Adding a His-tag cleavage step to the Cmr2:3 purification protocol has not affected the yield of pure protein after size exclusion chromatography, and initial screening of crystallization conditions have yielded crystal florets. Secondary crystallization trays will be set up with titrations of crystallization condition components, with the hope of sending reproducible crystals to be diffracted at the Stanford Synchrotron Radiation Lightsource.

Csm Subunit Expression and Purification

Although we were unsuccessful in purifying the Csm complex with our existing purification protocols, there is promise in greater subunit expression in TB media given the higher concentration of Csm1:4 heterodimer produced compared to growth in LB

media. The purification buffers have been adapted from existing purification protocols for the Type III-B Cmr complex, and there is also merit in optimization of salt and reducing agent concentrations as nucleic acid content is a prevalent issue in our existing Csm purification protocol. Another route of optimization is deciding which subunit to express in the pSATL vector, as we have not tried different combinations of tagged and untagged subunits. Codon-optimizing certain subunits that fail to express is also a viable future direction, as it has been successful in facilitating the consistent expression of Cmr2 and Cmr3.

Csm1:4 Purification and Crystallization

Following Csm complex purification protocols with T7 Express cells grown in TB has yielded enough protein to set up one 96-well crystallization screen per 6L of TB grown. Although some crystals have formed in numerous conditions, there was restrictions in the output of secondary crystallization trays to optimize the conditions of each crystallization screen hit as there is a very limited amount of purified protein produced per 6L of TB media.

The successful expression of the Csm1:4 heterodimer demonstrates that there is merit in co-transformation of subunits that are known to associate in the wild type Csm complex, with plans for other combinations of Csm subunits planned for the future.

BIBLIOGRAPHY

Arslan, Zihni, Veronica Hermanns, Reinhild Wurm, Rolf Wagner, and Umit Pul. "Detection and Characterization of Spacer Integration Intermediates in Type I-E CRISPR-Cas System." *Nucleic Acids Research* 42.12 (2014): 7884-893. Web. 11 Mar. 2017.

Babu, Mohan, Natalia Beloglazova, Robert Flick, Chris Graham, Tatiana Skarina, Boguslaw Nocek, Alla Gagarinova, Oxana Pogoutse, Greg Brown, Andrew Binkowski, Sadhna Phanse, Andrzej Joachimiak, Eugene V. Koonin, Alexei Savchenko, Andrew Emili, Jack Greenblatt, Aled M. Edwards, and Alexander F. Yakunin. "A Dual Function of the CRISPR-Cas System in Bacterial Antivirus Immunity and DNA Repair." *Molecular Microbiology* 79.2 (2010): 484-502. Web. 4 Mar. 2017.

Bai, Xiao-Chen, Greg Mcmullan, and Sjors H.w Scheres. "How Cryo-EM Is Revolutionizing Structural Biology." *Trends in Biochemical Sciences* 40.1 (2015): 49-57. Web. 6 Mar. 2017.

Barrangou, Rodolphe, and John Van Der Oost. "CRISPR-Cas Systems." *Springer* (2013): n. pag. Web. 14 Mar. 2017.

Beloglazova, N., G. Brown, M. D. Zimmerman, M. Proudfoot, K. S. Makarova, M. Kudritska, S. Kochinyan, S. Wang, M. Chruszcz, W. Minor, E. V. Koonin, A. M. Edwards, A. Savchenko, and A. F. Yakunin. "A Novel Family of Sequence-specific Endoribonucleases Associated with the Clustered Regularly Interspaced Short Palindromic Repeats." *Journal of Biological Chemistry* 283.29 (2008): 20361-0371. Web. 14 Mar. 2017.

Bikard, David, Asma Hatoum-Aslan, Daniel Mucida, and Luciano A. Marraffini. "CRISPR Interference Can Prevent Natural Transformation and Virulence Acquisition during In Vivo Bacterial Infection." *Cell Host & Microbe* 12.2 (2012): 177-86. Web. 9 Mar. 2017.

Blosser, Timothy R., Luuk Loeff, Edze R. Westra, Marnix Vlot, Tim Kunne, Malgorzata Sobota, Cees Dekker, Stan J.j. Brouns, and Chirlmin Joo. "Two Distinct DNA Binding Modes Guide Dual Roles of a CRISPR-Cas Protein Complex." *Molecular Cell* 58.1 (2015): 60-70. Web. 19 Mar. 2017.

Bobay, Louis-Marie, Marie Touchon, and Eduardo P. C. Rocha. "Manipulating or Superseding Host Recombination Functions: A Dilemma That Shapes Phage Evolvability." *PLoS Genetics* 9.9 (2013): n. pag. Web. 14 Mar. 2017.

Bolotin, Alexander, Benoit Quinquis, Alexei Sorokin, and Dusko S. Ehrlich. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin." *Microbiology* 151.8 (2005): 2551-561. Web. 14 Mar. 2017.

Brouns, S. J. J., M. M. Jore, M. Lundgren, E. R. Westra, R. J. H. Slijkhuis, A. P. L. Snijders, M. J. Dickman, K. S. Makarova, E. V. Koonin, and J. Van Der Oost. "Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes." *Science* 321.5891 (2008): 960-64. Web. 19 Mar. 2017.

Carte, J., N. T. Pfister, M. M. Compton, R. M. Terns, and M. P. Terns. "Binding and Cleavage of CRISPR RNA by Cas6." *Rna* 16.11 (2010): 2181-188. Web. 14 Mar. 2017.

Carte, J., R. Wang, H. Li, R. M. Terns, and M. P. Terns. "Cas6 Is an Endoribonuclease That Generates Guide RNAs for Invader Defense in Prokaryotes." *Genes & Development* 22.24 (2008): 3489-496. Web. 14 Mar. 2017.

Datsenko, Kirill A., Ksenia Pougach, Anton Tikhonov, Barry L. Wanner, Konstantin Severinov, and Ekaterina Semenova. "Molecular Memory of Prior Infections Activates the CRISPR/Cas Adaptive Bacterial Immunity System." *Nature Communications* 3 (2012): 945. Web. 4 Mar. 2017.

Deltcheva, Elitza, Krzysztof Chylinski, Cynthia M. Sharma, Karine Gonzales, Yanjie Chao, Zaid A. Pirzada, Maria R. Eckert, Jorg Vogel, and Emmanuelle Charpentier. "CRISPR RNA Maturation by Trans-encoded Small RNA and Host Factor RNase III." *Nature* 471.7340 (2011): 602-07. Web. 14 Mar. 2017.

Deng, Ling, Roger A. Garrett, Shiraz A. Shah, Xu Peng, and Qunxin She. "A Novel Interference Mechanism by a Type IIIB CRISPR-Cmr Module InSulfolobus." *Molecular Microbiology* 87.5 (2013): 1088-099. Web. 2 Mar. 2017.

Diez-Villasenor, Cesar, Noemi M. Guzman, Cristobal Almendros, Jesus Garcia-Martinez, and Francisco J.m. Mojica. "CRISPR-spacer Integration Reporter Plasmids Reveal Distinct Genuine Acquisition Specificities among CRISPR-Cas I-E Variants Of Escherichia Coli." *RNA Biology* 10.5 (2013): 792-802. Web. 19 Mar. 2017.

Elmore, Joshua R., Nolan F. Sheppard, Nancy Ramia, Trace Deighan, Hong Li, Rebecca M. Terns, and Michael P. Terns. "Bipartite Recognition of Target RNAs Activates DNA Cleavage by the Type III-B CRISPR-Cas System." *Genes & Development* 30.4 (2016): 447-59. Web. 4 Mar. 2017.

Erdmann, Susanne, and Roger A. Garrett. "Selective and Hyperactive Uptake of Foreign DNA by Adaptive Immune Systems of an Archaeon via Two Distinct Mechanisms." *Molecular Microbiology* 85.6 (2012): 1044-056. Web. 4 Mar. 2017.

Erdmann, Susanne, Shiraz A. Shah, and Roger A. Garrett. "SMV1 Virus-induced CRISPR Spacer Acquisition from the Conjugative Plasmid PMGB1 In Sulfolobus Solfataricus P2." *Biochemical Society Transactions* 41.6 (2013): 1449-458. Web. 12 Mar. 2017.

Estrella, Michael A., Fang-Ting Kuo, and Scott Bailey. "RNA-activated DNA Cleavage by the Type III-B CRISPR-Cas Effector Complex." *Genes & Development* 30.4 (2016): 460-70. Web. 2 Mar. 2017.

Fineran, P. C., M. J. H. Gerritzen, M. Suarez-Diez, T. Kunne, J. Boekhorst, S. A. F. T. Van Hijum, R. H. J. Staals, and S. J. J. Brouns. "Degenerate Target Sites Mediate Rapid Primed CRISPR Adaptation." *Proceedings of the National Academy of Sciences* 111.16 (2014): n. pag. Web. 7 Mar. 2017.

Garneau, Josiane E., Marie-Eve Dupuis, Manuela Villion, Dennis A. Romero, Rodolphe Barrangou, Patrick Boyaval, Christophe Fremaux, Philippe Horvath, Alfonso H. Magadan, and Sylvain Moineau. "The CRISPR/Cas Bacterial Immune System Cleaves Bacteriophage and Plasmid DNA." *Nature* 468.7320 (2010): 67-71. Web. 9 Mar. 2017.

Gasiunas, G., R. Barrangou, P. Horvath, and V. Siksnys. "Cas9-crRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria." *Proceedings of the National Academy of Sciences* 109.39 (2012): n. pag. Web. 2 Mar. 2017.

Goldberg, Gregory W., Wenyan Jiang, David Bikard, and Luciano A. Marraffini. "Conditional Tolerance of Temperate Phages via Transcription-dependent CRISPR-Cas Targeting." *Nature* 514.7524 (2014): 633-37. Web. 14 Mar. 2017.

Gophna, Uri, David M. Kristensen, Yuri I. Wolf, Ovidiu Popa, Christine Drevet, and Eugene V. Koonin. "No Evidence of Inhibition of Horizontal Gene Transfer by CRISPR-Cas on Evolutionary Timescales." *The ISME Journal* 9.9 (2015): 2021-027. Web. 5 Mar. 2017.

Hale, Caryn R., Alexis Cocozaki, Hong Li, Rebecca M. Terns, and Michael P. Terns. "Target RNA Capture and Cleavage by the Cmr Type III-B CRISPR-Cas Effector Complex." *Genes & Development* 28.21 (2014): 2432-443. Web. 4 Mar. 2017.

Hatoum-Aslan, A., P. Samai, I. Maniv, W. Jiang, and L. A. Marraffini. "A Ruler Protein in a Complex for Antiviral Defense Determines the Length of Small Interfering CRISPR RNAs." *Journal of Biological Chemistry* 288.39 (2013): 27888-7897. Web. 10 Mar. 2017.

He, Xiangjun, Chunlai Tan, Feng Wang, Yaofeng Wang, Rui Zhou, Dexuan Cui, Wenxing You, Hui Zhao, Jianwei Ren, and Bo Feng. "Knock-in of Large Reporter Genes in Human Cells via CRISPR/Cas9-induced Homology-dependent and Independent DNA Repair." *Nucleic Acids Research* 44.9 (2016): n. pag. Web. 14 Mar. 2017.

Hochstrasser, M. L., D. W. Taylor, P. Bhat, C. K. Guegler, S. H. Sternberg, E. Nogales, and J. A. Doudna. "CasA Mediates Cas3-catalyzed Target Degradation during CRISPR RNA-guided Interference." *Proceedings of the National Academy of Sciences* 111.18 (2014): 6618-623. Web. 13 Mar. 2017.

Horvath, P., and R. Barrangou. "CRISPR/Cas, the Immune System of Bacteria and Archaea." Science 327.5962 (2010): 167-70. Web. 19 Mar. 2017.

Ivancic-Bace, Ivana, Simon D. Cass, Stephen J. Wearne, and Edward L. Bolt. "Different Genome Stability Proteins Underpin Primed and Naive Adaptation In E. Coli CRISPR-Cas Immunity." *Nucleic Acids Research* 43.22 (2015): 10821-0830. Web. 14 Mar. 2017.

Jansen, Ruud., Jan. D. A. Van Embden, Wim. Gaastra, and Leo. M. Schouls. "Identification of Genes That Are Associated with DNA Repeats in Prokaryotes." *Molecular Microbiology* 43.6 (2002): 1565-575. Web. 14 Mar. 2017.

Jinek, M., F. Jiang, D. W. Taylor, S. H. Sternberg, E. Kaya, E. Ma, C. Anders, M. Hauer, K. Zhou, S. Lin, M. Kaplan, A. T. Iavarone, E. Charpentier, E. Nogales, and J. A. Doudna. "Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation." *Science* 343.6176 (2014): 1247997. Web. 16 Mar. 2017.

Kazlauskiene, Migle, Gintautas Tamulaitis, Georgij Kostiuk, Ceslovas Venclovas, and Virginijus Siksnys. "Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition." *Molecular Cell* 62.2 (2016): 295-306. Web. 13 Mar. 2017.

Kunin, Victor, Rotem Sorek, and Philip Hugenholtz. "Evolutionary Conservation of Sequence and Secondary Structures in CRISPR Repeats." *Genome Biology* 8.4 (2007): n. pag. Web. 19 Mar. 2017.

Levy, Asaf, Moran G. Goren, Ido Yosef, Oren Auster, Miriam Manor, Gil Amitai, Rotem Edgar, Udi Qimron, and Rotem Sorek. "CRISPR Adaptation Biases Explain

Preference for Acquisition of Foreign DNA." *Nature* 520.7548 (2015): 505-10. Web. 11 Mar. 2017.

Li, Hong. "Structural Principles of CRISPR RNA Processing." *Structure* 23.1 (2015): 13-20. Web. 15 Mar. 2017.

Li, M., R. Wang, and H. Xiang. "Haloarcula Hispanica CRISPR Authenticates PAM of a Target Sequence to Prime Discriminative Adaptation." *Nucleic Acids Research* 42.11 (2014): 7226-235. Web. 3 Mar. 2017.

Liu, T., Y. Li, X. Wang, Q. Ye, H. Li, Y. Liang, Q. She, and N. Peng. "Transcriptional Regulator-mediated Activation of Adaptation Genes Triggers CRISPR De Novo Spacer Acquisition." *Nucleic Acids Research* 43.2 (2015): 1044-055. Web. 15 Mar. 2017.

Makarova, K. S., Y. I. Wolf, and E. V. Koonin. "Comparative Genomics of Defense Systems in Archaea and Bacteria." *Nucleic Acids Research* 41.8 (2013): 4360-377. Web. 14 Mar. 2017.

Marraffini, L. A., and E. J. Sontheimer. "CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA." *Science* 322.5909 (2008): 1843-845. Web. 1 Mar. 2017.

Marraffini, Luciano A., and Erik J. Sontheimer. "Self versus Non-self Discrimination during CRISPR RNA-directed Immunity." *Nature* 463.7280 (2010): 568-71. Web. 4 Mar. 2017.

Mojica, F. J., C. Diez-Villasenor, J. Garcia-Martinez, and C. Almendros. "Short Motif Sequences Determine the Targets of the Prokaryotic CRISPR Defence System." *Microbiology* 155.3 (2009): 733-40. Web. 12 Mar. 2017.

Mulepati, S., and S. Bailey. "Structural and Biochemical Analysis of Nuclease Domain of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR)-associated Protein 3 (Cas3)." *Journal of Biological Chemistry* 286.36 (2011): 31896-1903. Web. 14 Mar. 2017.

Nunez, James K., Amy S. Y. Lee, Alan Engelman, and Jennifer A. Doudna. "Integrase-mediated Spacer Acquisition during CRISPR-Cas Adaptive Immunity." *Nature* 519.7542 (2015): 193-98. Web. 14 Mar. 2017.

Nunez, James K., Philip J. Kranzusch, Jonas Noeske, Addison V. Wright, Christopher W. Davies, and Jennifer A. Doudna. "Cas1-Cas2 Complex Formation Mediates Spacer Acquisition during CRISPR-Cas Adaptive Immunity." *Nature Structural & Molecular Biology* 21.6 (2014): 528-34. Web. 4 Mar. 2017.

Nunez, J.k., L.b. Harrington, P.j. Kranzusch, A.n. Engelman, and J.a. Doudna. "Crystal Structure the Escherichia Coli Cas1-Cas2 Complex Bound to Protospacer DNA." (2015): n. pag. Web. 14 Mar. 2017.

Osawa, Takuo, Hideko Inanaga, Chikara Sato, and Tomoyuki Numata. "Crystal Structure of the CRISPR-Cas RNA Silencing Cmr Complex Bound to a Target Analog." *Molecular Cell* 58.3 (2015): 418-30. Web. 28 Mar. 2017.

Paix, A., Y. Wang, H. E. Smith, C.-Y. S. Lee, D. Calidas, T. Lu, J. Smith, H. Schmidt, M. W. Krause, and G. Seydoux. "Scalable and Versatile Genome Editing Using Linear DNAs with Microhomology to Cas9 Sites in Caenorhabditis Elegans." *Genetics* 198.4 (2014): 1347-356. Web. 5 Mar. 2017.

Peng, W., M. Feng, X. Feng, Y. X. Liang, and Q. She. "An Archaeal CRISPR Type III-B System Exhibiting Distinctive RNA Targeting Features and Mediating Dual RNA

and DNA Interference." *Nucleic Acids Research* 43.1 (2014): 406-17. Web. 6 Mar. 2017.

Peranen, Johan, Marja Rikkonen, Marko Hyvonen, and Leevi Kaariainen. "T7 Vectors with a Modified T7lacPromoter for Expression of Proteins InEscherichia Coli." *Analytical Biochemistry* 236.2 (1996): 371-73. Web. 11 Apr. 2017.

Pougach, Ksenia, Ekaterina Semenova, Ekaterina Bogdanova, Kirill A. Datsenko, Marko Djordjevic, Barry L. Wanner, and Konstantin Severinov. "Transcription, Processing and Function of CRISPR Cassettes in Escherichia Coli." *Molecular Microbiology* 77.6 (2010): 1367-379. Web. 14 Mar. 2017.

Ramia, Nancy F., Michael Spilman, Li Tang, Yaming Shao, Joshua Elmore, Caryn Hale, Alexis Cocozaki, Nilakshee Bhattacharya, Rebecca M. Terns, Michael P. Terns, Hong Li, and Scott M. Stagg. "Essential Structural and Functional Roles of the Cmr4 Subunit in RNA Cleavage by the Cmr CRISPR-Cas Complex." *Cell Reports* 9.5 (2014): 1610-617. Web. 17 Mar. 2017.

Ran, F. Ann, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. "Genome Engineering Using the CRISPR-Cas9 System." *Nature Protocols* 8.11 (2013): 2281-308. Web. 11 Mar. 2017.

Rath, Devashish, Lina Amlinger, Archana Rath, and Magnus Lundgren. "The CRISPR-Cas Immune System: Biology, Mechanisms and Applications." *Biochimie* 117 (2015): 119-28. Web. 7 Mar. 2017.

Redding, Sy, Samuel H. Sternberg, Myles Marshall, Bryan Gibb, Prashant Bhat, ChantalÂ K. Guegler, Blake Wiedenheft, JenniferÂ A. Doudna, and Eric C.

Greene. "Surveillance and Processing of Foreign DNA by the Escherichia Coli CRISPR-Cas System." *Cell* 163.4 (2015): 854-65. Web. 15 Mar. 2017.

Reeks, Judith, James H. Naismith, and Malcolm F. White. "CRISPR Interference: A Structural Perspective." *Biochemical Journal* 453.2 (2013): 155-66. Web. 14 Mar. 2017.

Richardson, Christopher D., Graham J. Ray, Mark A. Dewitt, Gemma L. Curie, and Jacob E. Corn. "Enhancing Homology-directed Genome Editing by Catalytically Active and Inactive CRISPR-Cas9 Using Asymmetric Donor DNA." *Nature Biotechnology* 34.3 (2016): 339-44. Web. 14 Mar. 2017.

Samai, Poulami, Nora Pyenson, Wenyan Jiang, Gregory W. Goldberg, Asma Hatoum-Aslan, and Luciano A. Marraffini. "Co-transcriptional DNA and RNA Cleavage during Type III CRISPR-Cas Immunity." *Cell* 161.5 (2015): 1164-174. Web. 10 Mar. 2017.

Samson, Julie E., Alfonso H. Magadan, Mourad Sabri, and Sylvain Moineau. "Revenge of the Phages: Defeating Bacterial Defences." Nature Reviews Microbiology 11.10 (2013): 675-87. Web. 19 Mar. 2017.

Sashital, Dipali G., Blake Wiedenheft, and Jennifer A. Doudna. "Mechanism of Foreign DNA Selection in a Bacterial Adaptive Immune System." *Molecular Cell* 46.5 (2012): 606-15. Web. 14 Mar. 2017.

Savitskaya, Ekaterina, Ekaterina Semenova, Vladimir Dedkov, Anastasia Metlitskaya, and Konstantin Severinov. "High-throughput Analysis of Type I-E CRISPR/Cas Spacer Acquisition InE. Coli." *RNA Biology* 10.5 (2013): 716-25. Web. 12 Mar. 2017.

Schumann, Kathrin, Steven Lin, Eric Boyer, Dimitre R. Simeonov, Meena Subramaniam, Rachel E. Gate, Genevieve E. Haliburton, Chun J. Ye, Jeffrey A. Bluestone, Jennifer A. Doudna, and Alexander Marson. "Generation of Knock-in Primary Human T Cells Using Cas9 Ribonucleoproteins." *Proceedings of the National Academy of Sciences* 112.33 (2015): 10437-0442. Web. 14 Mar. 2017.

Semenova, E., M. M. Jore, K. A. Datsenko, A. Semenova, E. R. Westra, B. Wanner, J. Van Der Oost, S. J. J. Brouns, and K. Severinov. "Interference by Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) RNA Is Governed by a Seed Sequence." *Proceedings of the National Academy of Sciences* 108.25 (2011): 10098-0103. Web. 13 Mar. 2017.

Shao, Yaming, Alexis I. Cocozaki, Nancy F. Ramia, Rebecca M. Terns, Michael P. Terns, and Hong Li. "Structure of the Cmr2-Cmr3 Subcomplex of the Cmr RNA Silencing Complex." *Structure* 21.3 (2013): 376-84. Web. 19 Mar. 2017.

Shi, Junwei, Eric Wang, Joseph P. Milazzo, Zihua Wang, Justin B. Kinney, and Christopher R. Vakoc. "Discovery of Cancer Drug Targets by CRISPR-Cas9 Screening of Protein Domains." *Nature Biotechnology* 33.6 (2015): 661-67. Web. 14 Mar. 2017.

Sinkunas, Tomas, Giedrius Gasiunas, Christophe Fremaux, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. "Cas3 Is a Single-stranded DNA Nuclease and ATP-dependent Helicase in the CRISPR/Cas Immune System." *The EMBO Journal* 30.7 (2011): 1335-342. Web. 4 Mar. 2017.

Spilman, Michael, Alexis Cocozaki, Caryn Hale, Yaming Shao, Nancy Ramia, Rebeca Terns, Michael Terns, Hong Li, and Scott Stagg. "Structure of an RNA Silencing

Complex of the CRISPR-Cas Immune System." *Molecular Cell* 52.1 (2013): 146-52. Web. 13 Mar. 2017.

Staals, Raymond H.j., Yifan Zhu, David W. Taylor, Jack E. Kornfeld, Kundan Sharma, Arjan Barendregt, Jasper J. Koehorst, Marnix Vlot, Nirajan Neupane, Koen Varossieau, Keiko Sakamoto, Takehiro Suzuki, Naoshi Dohmae, Shigeyuki Yokoyama, Peter J. Schaap, Henning Urlaub, Albert J.r. Heck, Eva Nogales, Jennifer A. Doudna, Akeo Shinkai, and John Van Der Oost. "RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus Thermophilus." *Molecular Cell* 56.4 (2014): 518-30. Web. 1 Mar. 2017.

Staals, Raymond H.j., Yoshihiro Agari, Saori Maki-Yonekura, Yifan Zhu, David W. Taylor, Esther Van Duijn, Arjan Barendregt, Marnix Vlot, Jasper J. Koehorst, Keiko Sakamoto, Akiko Masuda, Naoshi Dohmae, Peter J. Schaap, Jennifer A. Doudna, Albert J.r. Heck, Koji Yonekura, John Van Der Oost, and Akeo Shinkai. "Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus Thermophilus." *Molecular Cell* 52.1 (2013): 135-45. Web. 14 Mar. 2017.

Sternberg, Samuel H., Sy Redding, Martin Jinek, Eric C. Greene, and Jennifer A. Doudna. "DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9." *Biophysical Journal* 106.2 (2014): n. pag. Web. 14 Mar. 2017.

Sternberg, Samuel H., Hagen Richter, Emmanuelle Charpentier, and Udi Qimron. "Adaptation in CRISPR-Cas Systems." *Molecular Cell* 61.6 (2016): 797-808. Web. 4 Mar. 2017.

Szczelkun, Mark D., Maria S. Tikhomirova, Tomas Sinkunas, Giedrius Gasiunas, Tautvydas Karvelis, Patrizia Pschera, Virginijus Siksnys, and Ralf Seidel. "Direct Observation of R-loop Formation by Single RNA-guided Cas9 and Cascade Effector Complexes." *Proceedings of the National Academy of Sciences* 111.27 (2014): 9798-803. Web. 12 Mar. 2017.

Tamulaitis, Gintautas, Ceslovas Venclovas, and Virginijus Siksnys. "Type III CRISPR-Cas Immunity: Major Differences Brushed Aside." *Trends in Microbiology* 25.1 (2017): 49-61. Web. 10 Apr. 2017.

Tamulaitis, Gintautas, Migle Kazlauskiene, Elena Manakova, Ceslovas Venclovas, Alison O. Nwokeoji, Mark J. Dickman, Philippe Horvath, and Virginijus Siksnys. "Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus Thermophilus." *Molecular Cell* 56.4 (2014): 506-17. Web. 4 Mar. 2017.

Taylor, D. W., Y. Zhu, R. H. J. Staals, J. E. Kornfeld, A. Shinkai, J. Van Der Oost, E. Nogales, and J. A. Doudna. "Structures of the CRISPR-Cmr Complex Reveal Mode of RNA Target Positioning." *Science* 348.6234 (2015): 581-85. Web. 2 Mar. 2017.

Terns, Rebecca M., and Michael P. Terns. "The RNA- and DNA-targeting CRISPR-Cas Immune Systems OfPyrococcus Furiosus." *Biochemical Society Transactions* 41.6 (2013): 1416-421. Web. 17 Apr. 2017.

Thompson, Rebecca F., Matt Walker, C. Alistair Siebert, Stephen P. Muench, and Neil A. Ranson. "An Introduction to Sample Preparation and Imaging by Cryo-electron Microscopy for Structural Biology." *Methods* 100 (2016): 3-15. Web. 4 Apr. 2017.

Wang, Jiuyu, Jiazhi Li, Hongtu Zhao, Gang Sheng, Min Wang, Maolu Yin, and Yanli Wang. "Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems." *Cell* 163.4 (2015): 840-53. Web. 6 Mar. 2017.

Wang, Ruiying, Gan Preamplume, Michael P. Terns, Rebecca M. Terns, and Hong Li. "Interaction of the Cas6 Riboendonuclease with CRISPR RNAs: Recognition and Cleavage." *Structure* 19.2 (2011): 257-64. Web. 1 Mar. 2017.

Wei, Yunzhou, Rebecca M. Terns, and Michael P. Terns. "Cas9 Function and Host Genome Sampling in Type II-A CRISPR-Cas Adaptation." *Genes & Development* 29.4 (2015): 356-61. Web. 12 Mar. 2017.

Westra, Edze R, Paul B G Van Erp, Tim Kunne, Bollen, Sander, Ekaterina Semenova, Konstantin Severinov, Shi Pey Wong, Raymond HJ Staals, Christel LC Seegers, Matthijs M. Jore, Willem M. De Vos, Remus T. Dame, Renko De Vries, Stan JJ Brouns, and John Van Der Oost. "CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3." Molecular Cell 46.5 (2012): 595-605. Web. 14 Mar. 2017.

Wiedenheft, Blake, Kaihong Zhou, Martin Jinek, Scott M. Coyle, Wendy Ma, and Jennifer A. Doudna. "Structural Basis for DNase Activity of a Conserved Protein Implicated in CRISPR-Mediated Genome Defense." *Structure* 17.6 (2009): 904-12. Web. 15 Mar. 2017.

Yang, Chi-Dung, Yen-Hua Chen, Hsi-Yuan Huang, Hsien-Da Huang, and Ching-Ping Tseng. "CRP Represses the CRISPR/Cas System InEscherichia Coli: Evidence That Endogenous CRISPR Spacers Impede Phage P1 Replication." *Molecular Microbiology* 92.5 (2014): 1072-091. Web. 12 Mar. 2017.

Yosef, I., M. G. Goren, and U. Qimron. "Proteins and DNA Elements Essential for the CRISPR Adaptation Process in Escherichia Coli." *Nucleic Acids Research* 40.12 (2012): 5569-576. Web. 15 Mar. 2017.

Zhang, Jing, Christophe Rouillon, Melina Kerou, Judith Reeks, Kim Brugger, Shirley Graham, Julia Reimann, Giuseppe Cannone, Huanting Liu, Sonja-Verena Albers, James H. Naismith, Laura Spagnolo, and Malcolm F. White. "Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity." *Molecular Cell* 45.3 (2012): 303-13. Web. 19 Mar. 2017.

CURRICULUM VITAE

Bobby Kwan
Phone: 443-538-6246
Email: bkwan1@jhu.edu
Birth: January 25, 1992. MedStar Harbor Hospital, Baltimore, MD

EDUCATION

The Johns Hopkins University Bloomberg School of Public Health
Master of Science (Sc.M.) in Biochemistry & Molecular Biology Aug 2015 – May 2017
Doctor of Philosophy (Ph.D.) in Biochemistry & Molecular Biology Aug 2017 –

University of Maryland, Baltimore County
Bachelor of Science in Biochemistry & Molecular Biology May 2010 – May 2014

RESEARCH EXPERIENCE

Graduate Researcher Jan 2016-Present
Johns Hopkins School of Public Health, Department of Biochemistry & Molecular Biology
Principal Investigator: Scott Bailey

- Project #1: crystallization of *Thermotoga maritima* Csm1:Csm4 heterodimer
- Project #2: crystallization of *Thermotoga maritima* Cmr2:Cmr3 heterodimer
- Project #3: purification of *Thermotoga maritima* Cmr complex
- Responsibilities include round-the-horn and site-directed mutagenesis, DNA gel electrophoresis and extraction/purification, chromatography-based protein purification

Laboratory Research Technician Aug 2014 – Aug 2015
UMB School of Medicine, Department of Biochemistry, Center for Biomolecular Therapeutics
Principal Investigator: David Weber

- Project #1: purification and crystallization of human psoriasin
- Project #2: purification and crystallization of tissue transglutaminase
- Project #3: purification of ubiquitin-like protease 1 and Taro proteins
- Project #4: molecular replacement and electron density refinement of hS100A1 and hS100A4
- Responsibilities include utilizing fast protein liquid chromatogaphy, SDS/native PAGE, protein crystallization, dynamic light scattering, and soluble expression level tests

Undergraduate Researcher Apr 2013 - July 2014
UMBC Department of Chemistry
Principal Investigator: Songon An

- Research interest: investigate the *de novo* biosynthesis of purine via purinosomes and their organization in breast cancer/HeLa cells
- Worked on site-directed mutagenesis of purinosome enzyme phosphorylation sites to analyze the roles of glycinamide ribonucleotide formyltransferase and formylglycinamidine synthetase in the purine biosynthesis pathway
- Responsibilities include DNA/plasmid amplification, upkeep of plasmid libraries, cell passage, PCR, site-directed mutagenesis, usage of a hemocytometer, gel electrophoresis

RELATED WORK EXPERIENCE

Undergraduate Teaching Assistant July 2013 - Aug 2013
UMBC Department of Biology
BIOL300L (Experimental Biology Laboratory) Instructor: Karen Whitworth

- Worked with other TAs to ensure each lab section operates smoothly, troubleshoot problems or questions that students may have while they perform the lab activities, and clean and handle accidents that may occur

Microbiology Laboratory Assistant Jan 2013 - May 2013
BIOL275L (Microbiology Laboratory) Instructor: Susan Gdovin

- Primary job duties include preparation of bacterial cultures and plates for each lab, set up lab equipment and growth media for each lab section, and working with professors and TAs to ensure labs operate smoothly

Computational Modeling Intern Jan 2013 - Apr 2013
US Food & Drug Administration
Director: John Karanian

- Primary duties include analysis of porcine CT scans and computationally develop 3D reconstructions of certain cardiovascular areas of interest, such as 3D reconstructions of the arteries to aid in development of stints for atherosclerosis patients by manually delineating the areas of the arteries in each CT scan slice

Chemistry Lab Technician Jul 2012 - Nov 2012
UMBC Department of Chemistry

- Primary job duties include sanitizing and distributing glassware/equipment to all academic labs, preparation of reagents and buffers, set up/dismantle each week's lab for the introductory/organic/analytical labs, handling and proper disposal of chemical waste, and working with professors and coworkers to plan and sort out lab sessions each week

LEADERSHIP EXPERIENCE

Johns Hopkins Graduate Consulting Club — Mar 2016 – Present
- Director of Communications — May 2016 – Present
- Biotech & Healthcare Case Competition 2016 (Leader of Team BioScolutions)

JHSPH ScM Student Representative — Oct 2016 – May 2017

UMBC Mens' Club Lacrosse — Aug 2010 – May 2014
- Vice President (Second Year) — Aug 2013 – May 2014
- Vice President (First Year) — Aug 2011 – May 2012

Johns Hopkins University Mens' Club Lacrosse — Aug 2015 – Present
- Captain — Jun 2016 – Present